DOCUMENT RESUME

ED 329 784 CE 057 381

AUTHOR Schuemer, Rudolf, Ed.
TITLE Evaluation Concepts and Practices in Selected Distance Education Institutions.
INSTITUTION FernUniversitat, Hagen (West Germany). Zentrales Inst. fur Fernstudienforschung Arbeitsbereich.
PUB DATE Jan 91
NOTE 236p.
PUB TYPE Collected Works - General (020) -- Reports - Research/Technical (143)

EDRS PRICE MF01/PC10 Plus Postage.
DESCRIPTORS Developed Nations; *Distance Education; Educational Quality; *Educational Research; *Evaluation Methods; Evaluation Research; Foreign Countries; Higher Education; Instructional Material Evaluation; Program Effectiveness; Program Evaluation; Student Evaluation; Universities
IDENTIFIERS Australia; Costa Rica; Great Britain; Hong Kong; India; Israel; Italy; New York; Norway; Venezuela; West Germany

ABSTRACT

This report contains, in addition to the introduction and preface, 13 papers written by individuals working in the field of evaluation who present the concepts and practices of evaluation at their own particular distance education institutions. The introduction (Schuemer) gives a short outline of the evaluation nomenclature and an overview of the contributors to the report. The papers are: "Orientation and Conceptualization of the Academic Quality Control Centre" (Alavarado et al.); "Evaluation at the FernUniversitat" (Bartels); "System Evaluation: Some Reflections on Current Tertiary Practices in Australia and Hong Kong" (Castro); "Evaluation at a Distance" (Ganor); "Evaluation at Indira Gandhi National Open University, New Delhi, India" (Koul); "Assessment at Empire State College--Strategies and Methods Used in Evaluating Distance Education" (Lehmann, Granger); "Evaluation of Danish Distance Education at University Level" (Lorentsen); "Institutional Evaluation of the National Open University of Venezuela (U.N.A.)--A Proposal" (De Machado, Machado); "Evaluation in University Distance Education" (Moehle, Gerhardt); "Evaluation Concepts and Methods in Distance Education at Humboldt University (HUB)" (Nerlich); "Evaluation Concepts and Practices at NKI, Norway" (Rekkedal); "Evaluation in the Distance Courses Established by the Department of Education of 'La Sapienza' University (Rome)" (Vertecchi); and "Evaluation at the British Open University" (Woodley). (YLB)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

ED329784

FernUniversität
The Distance Teaching University
of the Federal Republic of Germany

Z I F F
Zentrales Institut für Fernstudienforschung
The Institute for Research into Distance Education

Evaluation Concepts and Practice in selected Distance Education Institutions

Edited by

Rudolf Schuemer

BEST COPY AVAILABLE

Hagen: January 1991

U.S. DEPARTMENT OF EDUCATION
Office of Educational Research and Improvement
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

☑ This document has been reproduced as received from the person or organization originating it.

☐ Minor changes have been made to improve reproduction quality.

• Points of view or opinions stated in this document do not necessarily represent official OERI position or policy.

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY

D Holmberg

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)."

CE057381

Zentrales Institut für Fernstudienforschung (ZIFF)
FernUniversität - Gesamthochschule -
Postfach 940
D-5800 Hagen 1 (W)
Federal Republic of Germany

(c) FernUniversität - Gesamthochschule -: Hagen, January 1991

FernUniversität
The German Distance Teaching University

Zentrales Institut für Fernstudienforschung (ZIFF)
Institute for Research into Distance Education

Evaluation Concepts and Practice in Selected Distance Education Institutions

edited by

Rudolf Schuemer

with contributions by

J. Alavarado B., G. D'Agostino S. & G. Bolaños M., U.N.E.D., Costa Rica
J. Bartels, FernUniversität, Germany
Angela Castro, Baptist College, Hong Kong
Margalit Ganor, Open University of Israel (O.U.I.)
B.N. Koul, Indira Gandhi National Open University (IGNOU), New Delhi, India
T. Lehmann & D. Granger, Empire State College (ESC), U.S.A
Annette Lorentsen, University of Aalborg, Denmark
Arlette U. De Machado & L. Machado, National Open University of Venezuela (U.N.A)
H. Möhle & H.-P. Gerhardt, Karl-Marx-Universität Leipzig, Germany (ex-GDR)
B.P. Nerlich, Humboldt-Universität zu Berlin, Germany (ex-GDR)
T. Rekkedal, NKI, Norway
B. Vertecchi, "La Sapienza" University (Rome), Italy
A. Woodley, The Open University, U.K.

Hagen: FernUniversität / ZIFF, January 1991

Contents

Introduction and Preface
R. Schuemer, FernUniversität, Germany 1

Orientation and Conceptualization of the Academic Quality Control Centre
J. Alavarado B., G. D'Agostino S. & G. Bolaños M., Universidad Estatal a Distancia, Costa Rica 37

Evaluation at the FernUniversität
J. Bartels, FernUniversität, Germany 54

System evaluation: Some reflections on current tertiary practices in Australia and Hong Kong
A. Castro, Baptist College, Hong Kong 69

Evaluation at a distance
M. Ganor, The Open University of Israel, Israel 80

Evaluation at Indira Gandhi National Open University, New Delhi, India
B.N. Koul, Indira Gandhi National Open University, India 89

Assessment at Empire State College - Strategies and methods used in evaluating distance education
T. Lehmann & D. Granger, Empire State College, U.S.A 102

Evaluation of Danish distance education at university level
A. Lorentsen, Aalborg University, Denmark 123

Institutional evaluation of the National Open University of Venezuela (U.N.A) - A proposal -
A. U. De Machado & L. Machado, Universidad Nacional Abierta, Venezuela 144

Evaluation in university distance education
H. Möhle & H.-P. Gerhardt, Karl-Marx-Universität Leipzig, Germany 150

Evaluation concepts and methods in distance education at Humboldt University
B.P. Nerlich, Humboldt Universität zu Berlin, Germany 164

Evaluation concepts and practices at NKI, Norway
T. Rekkedal, NKI, Norway 175

Evaluation in the distance courses established by the Department of Education of "La Sapienza" University (Rome)
B. Vertecchi, "La Sapienza" University, Italy 193

Evaluation at the British Open University
A. Woodley, The Open University, U.K. 208

List of Contributors 223

Introduction and Preface

Rudolf Schuemer
Institute for Research into Distance Education (ZIFF)
FernUniversität
Federal Republic of Germany

1 Function and addressees of this report

This report is concerned with the evaluation concepts and evaluation practice at various distance education institutions.

The literature covering evaluation and evaluation research is so extensive that it is difficult to survey. There is no lack of textbook-style or programmatic representations - be they for (programme) evaluation in general (e.g. Cronbach et al 1980; Cronbach 1982; Rossi & Freeman 1989; Wottawa & Thierau 1990; Weiss 1972) or be they for evaluation in the field of 'education'[1)] (e.g. Hamilton et al 1977 or Stufflebeam et al 1971) or else in the field of 'distance education' or 'distance learning' (e.g. Thorpe 1988); one also usually finds sections on 'evaluation' in books about distance education in general (e.g. Holmberg 1985, ch. 6; 1989, ch. 4: on course evaluation among other things; or Keegan 1990, part IV: the stress being laid on system evaluation). There is equally no lack of reports about concrete evaluation studies in distance education (some of which are quoted in the contributions to this report). - Cf. also the bibliography of writings on distance education by Holmberg (1990), which includes also several writings on evaluation in distance education.

Apart from this there are also generally quite a few contributions on evaluation in the proceedings of congresses on distance education (so for instance in those of the 'International congress of open and distance teaching universities' in Madrid 1984 or those of the international distance education congress which recently took place in Caracas: 15th World Conference ICDE, Caracas Venezuela 1990).

This report is neither a texbook nor a treatise on the general principles of evaluation in distance education, nor is it a report on the results of single concrete inquiries into evaluation on distance education or on components of distance education. It is, rather, a report situated on a medium level of abstraction: the concepts and the practices of evaluation in different distance study institutions are

1) 'Programme evaluation' and 'evaluation' in the field 'education' are not to be regarded as seperated categories or categories which exclude one another; they can overlap - as can also be seen in book titles like that of Cronbach's (1982), for example: "Designing evaluation of educational and social programs".

described. Authors who have been working in the field of evaluation have been given the opportunity to present the concepts and practices of evaluation at their own particular institution.

This report is aimed chiefly at the following **addressees**:

- practicians in the field of evaluation in distance education: it is to be hoped that they will gain new ideas and impulses for the practice of evaluation in their own institution from the report, or rather, from the presentation of that which other institutions are doing in the field of evaluation.

- newcomers to the field of evaluation in distance education: they can gain a first impression of the variety and complexity of the scope of duties of "evaluation in distance education".

- those working in the field of distance education (for instance course or programme developers and those responsible for courses or programmes): a better knowledge of the intended purposes of evaluation can perhaps contribute to reducing any reservations or prejudices which may still be in existence. It would be nice if both parties - those responsible for the development of the courses and programmes and those responsible for the evaluation - were to see themselves as partners in striving towards the common aim: the creation of optimal teaching and learning conditions.

The report is also addressed to decision-makers, respectively those on the management or in the administrative positions at distance study institutions, as a form of plea for more openness: for evaluation can only achieve something and contribute towards the improvement of "products" (teaching material, for instance) and "processes/course of progress" (for instance two-way communication) if the management of the respective institution is open to the necessities of evaluation, and is prepared to come to an agreement about the results of evaluation research, even when - or despite the fact that - those results may lead to the conclusion that not everything in the institution corresponds to the ideal situation. Evaluation remains more or less ineffectual if its results are not - at least internally - published and discussed.[2)]

The purpose of this introduction is to give:
- a short outline of the evaluation nomenclature (sections 2 – 3) and to give
- a first orientation regarding the contributions to this report (sections 4 – 6).

2) on this point Rekkedal writes in his contribution: "openness of institutional results is necessary for increasing quality."

2 Regarding the terms "evaluation" und "evaluation research"

There do not seem to be any generally accepted definitions for "evaluation" or "evaluation research" (some examples of definitions of and statements about evaluation are to be found in Box 1 on the following page; the statements there refer partially to programme evaluation, partially to evaluation in the field of education).

In a recently published textbook on evaluation (Wottawa & Thierau 1990, p. 9), a reference is made to Franklin & Trasher (1976, p. 20): "To say that there are as many definitions as there are evaluators is not far from accurate."

Rutman (1980, p. 17) is of similar opinion: "The term 'program evaluation' does not have a standardized and commonly accepted meaning. Rather, there are widely different interpretations of the term."

Attempts at reaching a closer definition of evaluation and evaluation research by differentiation were also unsuccessful or else have not been generally accepted. Suchman (1967), for instance, suggested differentiating between 'evaluation' and 'evaluation research'. Here '*evaluation*' denotes the process of assessing the worth of a product, a process or a programme, and does not necessarily contain systematic procedures or data-based evidence for supporting a judgement. '*Evaluation research*', on the other hand, denotes the explicit usage of scientific research methods and techniques for the means of carrying out a valuation.[3] Many authors, however, use both terms ('evaluation' and 'evaluation research') implicitly or explicitly synonymously (like, for instance, Rossi & Freeman 1989, p. 18).

3) Further differentiation is to be found in Guba (1969) or Abramson (1979) - cf. also the discussion in Wittmann (1985, section 3.1, p. 19ff).

Box 1: Examples for definitions of evaluation and statements on evaluation
"Evaluation is an *observed value* compared to some *standard*." Stake (1983, p. 291, basing on M. Scriven)
"Evaluation is the marshalling of information for the purpose of improving decisions." Thompson (1975, p. 8) on programme evaluation
"Evaluation is the process of delineating, collecting and providing information useful for judging decision alternatives." "Evaluation is the determination of the worth of a thing. It includes information for use in judging the worth of a program, product, procedure or objective or the potential utility of alternative approaches to attain specific objectives." Wentling & Lawson 1975 (quoted from Wittmann 1985, p. 20)
"...program evaluation refers to the use of research methods to measure the effectiveness of operating programs." Rutman (1980, p. 17) on programme evaluation
"The purpose of evaluation research is to measure the effects of a program against the goals it set out to accomplish as a means of contributing to subsequent decision making about the program and improving future programming." Carol H. Weiss (1972, p. 4) on the aim/purpose of programme evaluation
"Evaluation research is the systematic application of social research procedures for assessing the conceptualization, design, implementation, and utility of social intervention programs." Rossi & Freeman (1989, p. 18) as a definition of (programme) evaluation.
"By the term *evaluation*, we mean systematic examination of events occuring in and consequent on a contemporary program - an examination conducted to assist in improving this program and other programs having the same general purpose." Cronbach et al 1980, p. 14
"...we may define evaluation broadly as the *collection and use of information to make decisions about an educational program*. This program may be a set of instructional materials distributed nationally, the instructional activities of a single school, or the educational experiences of a single pupil. Many types of decision are to be made, and many varieties of information are useful. It becomes immediately apparent that evaluation is a diversified activity and that no one set of principles will suffice for all situations." Cronbach 1983, pp. 101-102
"Educational evaluation is the process of delineating, obtaining and providing useful information for judging decision alternatives." Stufflebeam et al 1971, p. 43 (quoted from McCormick & James 1989, p. 172)
"Evaluation is the process of conceiving, obtaining and communicating information for the guidance of educational decision-making with regard to a specified programme." MacDonald 1973, pp. 1-2, (quoted from McCormick & James 1989, p. 172)
"Curriculum evaluation is the process of delineating, obtaining and providing information useful for making decisions and jugdments about curricula." Davis 1981, p. 49 (quoted from McCormick & James 1989, p. 172)

Neither have attempts to order the approaches to evaluation in the form of taxonomies gained general acceptance. House (1980, ch. 2, especially fig. 1, p. 23), for instance, suggested a 'taxonomy of major evaluation approaches', in which different models are distinguished (cf. Box 2, next page).

According to Wottawa & Thierau (1990, p. 9), despite the lack of a generally accepted definition, some general characteristics of scientific evaluation can be given on which there is more or less agreement:

(1) Evaluation has something to do with valuation. Evaluation serves to help in planning and deciding, and thus has something to do with assessment and valuation of alternative ways of acting.

(2) Evaluation is oriented towards aims and purposes. It primarily has the aim of checking practical measures, of improving them or of making decisions concerning them.

(3) Evaluation measures should reflect the current state of scientific techniques and research methods.

As far as the third characteristic of scientific evaluation named by Wottawa & Thierau (the adaptation to the current state of scientific techniques) is concerned, one can say: the formulation is suited to disguise the fact that there definitely are differences of opinion as to what can be considered in each case as the most suitable methods and approaches or rather paradigms:

Glass & Ellett (1980, p. 211; quoted from Wittmann 1985, p. 40) write: "Evaluation - more than any science - is what people say it is, and people currently are saying it is many different things. Evaluation is a set of theoretical and practical activities without a widely accepted paradigm. Few people agree on the best way to evaluate."

Box 2: "A taxonomy of major evaluation approaches" (according to House 1980, ch. 2, fig. 1, p. 23)

*model**: systems analysis; *major audiences or reference groups*: economists, managers;
assumes consensus on: goals, known cause and effect, quantified variables;
methodology: amongst other things linear programming, planned variation, cost-benefit analysis;
outcome: efficency;
typical questions: Are the expected effects achieved? Can the effects be achieved more economically? What are the *most* efficient programs?

model: behavioral objectives; *audiences*: managers, psychologists;
consensus on: prespecified objectives, quantified outcome variables;
methodology: behavioral objectives, achievement tests;
outcome: productivity, accountability;
questions: Is the program achieving the objectives? Is the program producing?

model: decision making; *audiences*: decision-makers, especially administrators;
consensus on: general goals, criteria;
methodology: surveys, questionnaires, interviews, natural variation;
outcome: effectiveness, quality control;
questions: Is the program effective? What parts are effective?

model: goal free; *audiences*: consumers;
consensus on: consequences, criteria;
methodology: bias control, logical analysis, modus operandi;
outcome: consumer choice, social utility;
questions: What are *all* the effects?

model: art criticism; *audiences*: connoisseurs, consumers;
consensus on: critics, standards;
methodology: critical review;
outcome: improved standards, heightened awareness;
questions: Would a critic approve this program? Is the audience's appreciation increased?

model: professional review; *audiences*: professionals, public;
consensus on: criteria, panel, procedures;
methodology: review by panel, self study;
outcome: professional acceptance;
questions: How would professionals rate this program?

model: quasi-legal; *audiences*: jury;
consensus on: procedures and judges;
methodology: quasi-legal procedures;
outcome: resolution;
questions: What are the arguments for and against the program?

model: case study; *audiences*: client, practitioners;
consensus on: negotiations, activities;
methodology: case studies, interviews, observations;
outcome: understanding diversity;
questions: What does the program look like to different people?

* The viewpoints according to which things are ordered are printed in italics.

Thus there are, for instance, different opinions on the usage of more quantitative or else more qualitative methods/procedures, or on the appropriateness of approaches belonging more to the sphere of natural sciences, aiming at "explanation" as opposed to anthropological approaches aiming at "understanding" (found especially in the field of educational evaluation).
Such a disagreement is to be seen, for instance, in the following paragraph quoted from Patton (1980, p. 19):

> "Evaluation research is dominated by the largely unquestioned, natural science paradigm of hypothetico-deductive methodology. This dominant paradigm assumes quantitative measurement, experimental design, and multivariate, parametric statistical analysis to be the epitome of 'good' science. ( ) By way of contrast, the alternative to the dominant hypothetico-deductive paradigm is derived from the tradition of anthropological field studies. Using the techniques of in-depth, open-ended interviewing and personal observation, the alternative paradigm relies on qualitative data, holistic analysis, and detailed description derived from close contact with the targets of study. The hypothetico-deductive, natural science paradigm aims at prediction of social phenomena; the holistic-inductive, anthropological paradigm aims at understanding of social phenomena."
> (Patton 1980, p. 19)

A discussion on the "conflicting ideals for evaluation" (as they are termed in contrasts such as "humanistic vs. scientistic ideals" or "qualitative vs. quantitative approaches") is also to be found in Cronbach's work (1982, especially ch. 1). Cronbach pleads for a conscious effort to overcome such dichotomies; in the same book he writes later (p. 301):

> "Some of the debates regarding evaluation have encouraged the mistaken impression that objective, quantitative, focussed methods are incompatible with humanistic, qualitative, wide-band inquiry. In fact, the two should be working hand in hand."
> Cronbach 1982, p. 301

Similarly Patton, in two more recent books (1987, 1990), pleads for more pragmatism (cf. Box 3 on the following page).

(As regards the debate about qualitative and quantitative methods and approaches see also Reichardt & Cook 1979; cf. also Wittmann 1985, section 4.5, p. 180ff. - This problem is also considered in more detail in Lorentsen's paper in this report.)

Box 3: Patton on the "qualitative vs. quantitative" debate

"In the early literature on evaluation methods the debate between qualitative and quantitative methodologies was often strident. In recent years the debate has softened. A consensus has gradually emerged that the important challenge is to match appropriate methods to evaluation questions and issues, not to advocate universally any single methodological approach for all evaluation situations".
Patton 1987, pp. 168-169

"In short, evaluation has moved into a period of methodological diversity with a focus on methodological appropriateness".
Patton 1987, p. 169

"Because qualitative and quantitative methods involve differing strengths and weaknesses, they constitute alternative, but not mutually exclusive, strategies for research. Both qualitative and quantitative data can be collected in the same study."
Patton (1990, p. 14)

"I prefer pragmatism ... to one-sided paradigm allegiance".
Patton (1990, p. 38)

and under the heading "Pragmatism and a paradigm of choices" he writes:

"Rather than believing that one must choose to align with one paradigm or the other, I advocate a paradigm of choices. A paradigm of choices rejects methodological orthodoxy in favor of *methodological appropriateness* as the primary criterion for judging methodological quality. The issue then becomes not whether one has uniformly adhered to prescribed canons of either logical-positivism or phenomenology but whether one has made sensible methods decisions given the purpose of the inquiry, the questions being investigated, and the resources available. The paradigm of choices recognizes that different methods are appropriate for different situations."
Patton (1990, p. 39)

* The same book (Patton 1987, ch. 2) contains a list of situations and purposes, for which qualitative methods / procedures are especially suitable (e.g. for formative evaluation).

Regarding (1) and (2) in Wottawa & Thierau above (i.e., regarding the valuation and the aim- respectively purpose-orientation of evaluation): as regards the evaluation in educational institutions, one might add: evaluation or the determination of "value" is related to the aims of the educational institution in question. (And these, in turn, can have something to do with the aims of society as a whole or of the state.)

Educational institutions do not, however, only differ in their ways of procedure and forms of organization, but also in their aims (or at least in the way in which they emphasize those aims). If only for this reason the approaches to evaluation and the emphasis put on different evaluation subjects-matters will differ in different institutions.

3 Regarding evaluation in distance education

The attempts to find a generally agreed definition of evaluation have not been more successful in distance education than in other fields of evaluation.

The situation is made no easier by the fact that evaluation in distance education can refer to very different fields and levels. Thus Holmberg (1989, p. 170) refers to the fact that "the term 'evaluation' denotes different things in different contexts. Sometimes it refers to the assessment of students for the purpose of awarding marks, sometimes to the judgment of complete educational systems."

For the purposes of this report, Mary Thorpe's definition of evaluation (1988, p. 5) can - despite all the named difficulties - serve as a basis in the area of open and distance learning:

> "Evaluation is the collection, analysis and interpretation of information about any aspect of a programme of education and training, as part of a recognised process of judging its effectiveness, its efficiency and any other outcomes it may have."
> M. Thorpe (1988, p. 5)

Thorpe stresses the 'inclusiveness' of her definition; she expressly includes 'monitoring', too, as well as differing forms of evaluation (she mentions amongst others, formative, summative, continuous, 'illuminative' und 'goal-based' evaluation[4]).

She also points out, however, the very differing approaches to evaluation - also in particular in the field of 'education' (see the detailed quotation in Box 4 on the following page); she pleads (similarly to Cronbach 1982 or Patton 1987, 1990 with regard to quantitative vs. qualitative approaches - see above) for a pragmatic viewpoint:

> "...there is no one right model, and ... we can and should choose one or a combination of approaches appropriate to the purposes and resources available to us at the time.
> The procedures of an evaluation should be designed therefore to satisfy within a particular timescale the local purposes for which it was set up, which will vary according to the course or programme being evaluated."
> (M. Thorpe 1988, p. 9)

---

4) *Formative Evaluation* serves to supply information which can be used *during* the development of a programme or course. *Summative Evaluation* furnishes information about how well the *finished* programme or the finished course works. Cf. on this point also Thorpe (1988, pp. 9-13) or Box 6 below. - Regarding 'continuous evaluation' see amongst others the paper by Bartels. Re. 'illuminative evaluation' see amongst other things Box 4.

Box 4: M. Thorpe (1988, pp. 8-9) about different approaches to educational evaluation

"The development of different approaches to educational evaluation traces the effort to provide data about learning and processes of teaching which are richer and more informative than those provided by psychometric testing alone.

Tyler's 'Basic principles of Curriculum and Instruction' (1949) was one of the earliest benchmarks here. It stressed instructional objectives (preferably expressed in behavioral terms) as the measure of student progress and thereby the worth of the programme. One of the limitations of this approach was its narrow focus on quantitative assessment of student achievement, and its over-simplification of the school process as the context in which most of this evaluation occured. Subsequent approaches have picked up one or more of the perspectives which it ignores. Some evaluators have stressed the importance of the subjective accounts of practitioners and learners as well as findings and seek to promote understanding of the programme, rather than to measure outcomes.

Others have linked evaluation very closely to decision making within the institution, focussing on specifiying objectives and reviewing performance. This approach has also been criticised for its inward looking orientation and for an over-emphasis on intended effects. Evaluation should also explore outcomes in areas which the policy makers have ignored, or which are important to others. The effects of a programme may have wider and different reverberations than those predicted by the staff who implement it. One of the roles of evaluation is to remove the blinkers of those who are too close to a project to see some of its less obvious effects and implications.

These developments in evaluation generated much debate within the literature on educational evaluation about what should be its proper tasks and purposes. These arguments came to a head (in the U.K. at least) in a paper which drew together many similar approaches under the title of 'illuminative evaluation'. The authors of the paper, Parlett and Hamilton, challenged all approaches to educational evaluation which seek to bring research principles from the natural sciences into an arena of social interaction as in learning and teaching. Their charge against experimental controls, psychometric testing, the statistical manipulation of highly complex variables and dependence on quantitative assessment of learning, is that these methods are inappropriate for their subject matter and a missed opportunity. They ignore the social values which influence action within institutions of education and training and fail to illuminate the meaning of processes and events from the point of view of the key participants. They often remain remote from the decision making and development within institutions.

This critique of what Parlett and Hamilton refer to as the 'agricultural-botany paradigm' leads in to their outline of illuminative evaluation, which gives a greater role and status to qualitative evaluation. This offers a means of discovering competing interests among the various actors involved in institutional provision of education and training and of comprehending the complexity of the learning process.

I have included this reference to the differences of approach which exist within the field of evaluation not only to point out that there is no consensus around a 'best model' among the professionals, but also to hint at the variety of approaches and methods which are available. The aim is not to pick the winner from a league table but to emphasize that this variety exists, that there is no one right model, and that we can and should choose one or a combination of approaches appropriate to the purposes and resources available to us at the time.

The procedures of an evaluation should be designed therefore to satisfy within a particular timescale the local purposes for which it was set up, which will vary according to the course or programme being evaluated."

She goes on to add that evaluation in general covers the components 'data collection, analysis, and interpretation' and is 'a planned activity, capable of being made public'. On the latter she writes (p. 6):

> "What distinguishes evaluation from other related activities is not just that judgments and opinions are made, but that they are seen to be made. *Evaluation is essentially a public act - not necessarily published or publishable, but open to inspection by an outsider and therefore capable of being made public.*"
> (M. Thorpe 1988, p. 6)

This aspect can also be interpreted to mean that evaluation contributes or should contribute towards transparency of decisions and courses of action to be taken.

M. Thorpe goes on to stress the fact that not only the intended effects, but also the *un*intended effects of a measure can be made the object of an evaluation (p. 7): "The scope of evaluation is not limited to checks on the overt goals of a programme but can encompass any of its aspect and their effects."

M. Thorpe also continues to emphasize - similarly to Holmberg -, that evaluation can serve very different purposes, and gives a list of examples (see Box 5), noting however that (p. 7) "This is not a comprehensive list of purposes, nor are the individual items mutually exclusive."

Box 5: 'list of purposes of evaluation' in the field of open & distance learning (according to M. Thorpe 1988, p. 7)

(a) measurement of achievement of objectives of a programme as a whole;
(b) judging the effectiveness of courses or materials;
(c) finding out what the inputs into a programme were - number of staff, number and content of contact hours, time spent by the learner, and so on;
(d) 'mapping' the perceptions of different participants - learners, tutors, trainers,managers etc;
(e) exploring the comparative effectiveness of different ways of providing the same service;
(f) finding out any unintended effects of a programme, whether on learners, clients or open learning staff;
(g) regular feedback on progress towards meeting programme goals;
(h) finding out the kinds of help learners need at different stages;
(i) exploring the factors which appear to affect the outcomes of a programme or service.

At another point in the book (p. 198ff.) she differentiates between an approach which is aimed at 'mere quality control', and a 'developmental approach' (with stress on evaluation as a means towards continuous development and improving of teaching and learning). She writes (p. 202):

> "The danger of evaluation ... is that it is used narrowly as either a tool for personal or institutional justification or as a means of managerial control. If this is to be avoided, practitioners need to develop an alternative model ... . It takes a model to kick out a model; that is we cannot expect that a developmental approach to evaluation will come about merely by pointing out the shortcomings of a narrowly instrumental or control oriented approach."
> (Thorpe 1988, p. 202)

and slightly further on (pp. 202-203):

> "... I take it as axiomatic that evaluation is for development - of learning and the quality of learners' experiences, of service and of the quality of the curriculum the institution offers ..."
> (Thorpe 1988, pp. 202-203)

Here most of those involved in 'educational evaluation' will probably agree: evaluation should not end merely in a comparison of what should have been achieved and what actually has been achieved and in uncovering weak points (diagnosis aspect) - important as this may really be as a precondition for improvements; it should indicate as far as possible ways towards improving materials and prodecures and of increasing efficiency (taking into account the costs if and when necessary).

Following these lines and using a 'development-research continuum' ('pure development' - 'formative evaluation' - 'summative evaluation' - 'pure research') one might say that evaluation in distance education - at least when it is put into practice as internal evaluation - generally has a more formative than summative tendency (that is to say, it tends more towards 'development' than towards 'pure research'; - cf. also the continuum for the classification of formative and summative evaluation in Box 6, quoted from Borich 1974, pp. 28-29). It is also devoted to the constant (principally never-ending) attempts at improving the offer of courses, the services and the procedures.

Box 6: Continuum for the classification of formative and summative evaluation

In Borich (1974, pp. 28-29), a continumum is sketched out from *'pure development'* to *'pure evaluation'* via *planner*, *designer*, *developer*, *formative evaluator*, *summative evaluator*, *researcher*. The roles in this are described as follows (p. 29):

Planner:	organizes and marshals materials and personnel for the development task.
Designer:	writes content and performance specifications for the program or product.
Developer:	constructs product or program that meets content and performance specifications.
Formative evaluator:	assesses product or program components for the purpose of revision and modification.
Summative evaluator:	assesses product or program *in toto* for the purpose of adoption or continuation.
Researcher:	formulates and tests hypotheses concerning the program or product *vis a vis* theoretical framework.

Some typical tasks of evaluation in distance education

For the purpose of this report on evaluation in distance education the terms "evaluation" and "evaluation research" are to be interpreted as widely as possible - mainly due to the fact that the terms are interpreted differently in the various papers and also because the people or departments concerned with evaluation in the various different institutions have fields of work which are defined with differing degrees of broadness.

Some typical *tasks of evaluation* in distance education could be, for instance:[5)]

- 'monitoring'[6)] (assessing the degree of congruence between the plan for providing services and treatments and the ways they actually are provided): for instance monitoring/supervision of internal administration procedures and control of schedules; monitoring the mailing of the study materials or the processing of submitted assignments (correction and commenting; including the turn-around time); implementation control as regards the tutorials; etc.

- Analysis of the structure of applicants and participants and comparison of this structure with that of the target population: if, for instance, an educational programme/course is more emphatically aimed at underprivileged groups, one

5) The categories of tasks/fields of duties neither exclude one another nor is the list complete.

6) analoguous to the monitoring of the management and realization of "programmes" - cf. for instance Rossi & Freeman (1989, ch. 4: Program monitoring for evaluation, pp. 167-223). - 'Monitoring' can be considered narrowly in a more technical sense; it can, however, also be seen more widely and then includes 'quality control'.

can check to what extent these groups are represented amongst the applicants and participants. (It is counted among the aims of some of the "Open Universities", for example, to approach educationally as yet underprivileged groups more than before. - Cf. for instance the papers by Koul or Woodley.)

- Registering and assessment of the performance of the learners/students ('student assessment'):

  One may differ in opinion as to whether or not 'student assessment' *per se* belongs to the tasks of evaluation in distance study[7]. It is, however, hardly to be disputed that student assessment can offer important hints for course development or the improvement of courses (in the wider sense).

  For instance, the determination of the *previous knowledge* of the learners – especially at institutions with "open" admittance, that is admittance which is not dependent upon specific school qualifications; or with a heterogeneous target population – can give important hints as to what can be taken for granted in a course and what needs to be explained completely in all detail. (If necessary an offer of courses basing on different levels of basic knowledge can be worked out; cf. for instance, the paper by Vertecchi).

  The analysis of the *performance* of the learners (including examination results) can also be put to use for the purpose of evaluating the courses or also the system as a whole and for revising these: noticeably worse average performance in specific course units can, for instance, indicate didactic shortcomings in those units; similarly, bad final marks can hint at didactic shortcomings of a course and/or of the tutoring for the course.

  A low success rate in examinations can likewise be an indicator for shortcomings of programmes/curricula or the system as whole, etc.

  It is also absolutely clear that student assessment in institutions with a strong emphasis on 'individualized education' (like, for instance, the Empire State College - cf. the paper by Lehmann & Granger) is of central importance; it is that which first and foremost creates the prerequisites for the learner (and his/her tutor/counsellor), to plan purposefully the next steps in his/her individual course of studies and be able to put them into action.

  But even in cases where student assessment in an institution is not part of the job of those engaged in evaluation, pragmatic considerations speak in favour of the evaluators taking an interest in the field of assessment: due to their professional training they usually have good prerequisites (knowledge of test

7) On this theme see, for instance, M. Thorpe (1988, p. 6): " One of the terms evaluation is *not* synonymous with is assessment, which is the procedure of assigning value to the learning achieved during and at the end of a course. Examination results and continuous assessment are of course very relevant data for most evaluative purposes because they are a measure of outcomes. Like most other indicators, however, they do not so much provide the answers as suggest the need for further information and explanation."

analysis and test construction) for helping the specialists in matters of course content and didacticians in matters concerning the development of performance tests and examination questions.

- Evaluation of teaching materials and media: this stretches from the evaluation of single course units (or even only sections of these) to whole courses or even programmes or curricula (the latter two understood here to be a combination of different courses, each of which is composed of several units); in multi-media offers the evaluation may refer to the whole "package" or else only one single medium (e.g. the study letter as chief medium or the audio cassettes or diskette as a supporting medium). The evaluation can be carried out by questioning "experts" (e.g. colleagues, didacticians, media specialists) or by questioning the learners, and can be either more or less formalized. It can be restricted to the learning materials or can include the student services (see below).

- Evaluation of the student services (tutoring/counselling, including the correction of and commenting on assignments):

  Typical aspects are things like: 'turn-around time' for the correction of and comments on assignments submitted by students; detailedness and quality of the comments; reliability of the marking or performance ratings; friendliness towards the learners (in correspondence or telephone enquiries, for instance).

  If the offer of an institution also includes face-to-face tutorials, the behavior of the tutors shown there as well as the opinion of the students about them, are also to be included in the evaluation. (Video films of tutorial sessions, for instance, can be made and then later discussed with those involved; the video films can also be used in tutor training. - Cf. the paper by M. Ganor).

  A content analysis of the communication between learners and tutors can, in addition, offer hints at shortcomings in the contents and at comprehension difficulties which might be encountered in the course materials.

  The student services may be considered or evaluated together with the learning materials of a course or a programme or else in isolation.

- Studies on specific questions: for instance, the learner's access to certain media, respectively his/her possession of such media (like, for example, PCs or video recorders); time-budget inquiries amongst the learners (or specific target groups among the learners). Such inquiries may not really belong to evaluation in the strict sense; they are, however, relevant for the planning and developing of courses or programmes, or else when one needs to ascertain whether or not a certain course offer does not meet up with the learner's situation.

- Evaluation of the system as a whole: as regards, for example, the allocation of resources (finances, personnel), the number of those completing the course, respectively the drop-out rate[8] or the cost-effectiveness[9].

  The chances of the graduates on the labor market, as well as the opinion of potential employers about the institution and its graduates can be important indicators for the system evaluation.

  The contribution made by an institution to superior educational or overall political aims of society (contributions to the overall economic development, for instance, or to social innovation etc.) may also be the subject of evaluation.

From the fields of work named as examples above one can also clearly see that many of the evaluation activities are linked to the aims of the individual institutions. That is: that which is evaluated (and, possibly, how it is evaluated) depends to quite an extent upon the aims of the respective institution.

An institution like the Indira Ghandi National Open University, for example, which, parallel to its educational mission as a university also has other duties (such as, for instance, the promotion and supervision of distance studies in India), has to take into account a wider spectrum of subject-matter to be evaluated and of evaluation criteria than institutions which have "only" an educational mission (cf. the paper by Koul).

Or: an institution such as the Empire State College (ESC) with a stronger stress on 'individualized education' and as wide as possible a regard for individual study aims will lay the emphasis on points that are different from those emphasized by an institution like, for example, the FernUniversität with its relatively rigid curricular offers (emphasis on student assessment vs. emphasis on course evaluation; cf. the papers by Lehmann & Granger and Bartels).

Or: An institution which is committed to a 'communicative distance education concept' may tend to use other methods and approaches - for instance qualitative ones, aiming more at understanding than at explanation - than an institution which is committed to the 'knowledge distributing distance teaching concept' (cf. the paper by Lorentsen).

---

8) Regarding 'drop out' in distance education see, amongst others, Bååth (1984), Fritsch (1990), Peters (1988) or Schuemer & Ströhlein (1990) and the literature cited there. - Apart from 'drop out' other categories of the course of studies - like, for example, the change to a different programme/curriculum or to a different educational institution - can be relevant for evaluation.

9) On the aspects of costs in the evaluation of distance study systems see, amongst others, also Rumble (1983); Keegan 1990, ch. 10 or Holmberg (1989, Ch. 4, pp. 190-203). - Regarding the methodics of inquiries into effectivity (or cost-effectiveness) in the educational system see, amongst others, Weiß (1975).

Regarding the methods used in evaluation

The methods used in evaluation for data collection and analysis cover the whole range of methods of empirical social research[10] - like amongst others:

- written and oral (resp. telephone) questioning (in the case of written questioning mostly in a standardized way; with telephone or face-to-face interviews often also in a non-standardized form); individual and group interviews.
- Test procedures to assess the level of knowledge (e.g. of the previous knowledge) and the performance (e.g. standardized criterion-oriented achievement tests).
- Experiment and quasi-experiment (e.g. to examine the effectiveness of alternative forms of a course offer or of tutoring).
- Observation (e.g. in face-to-face tutorials).
- Content analysis (e.g. for written communication between learner and tutor).
- Cost-effectiveness or cost-benefit analysis (e.g. for individual programmes or also for the institution as a whole).

Which methods and procedures are used depends first and foremost, of course, on the respective subject of the evaluation inquiry (and the chosen forms of operationalization). Some criteria might be recorded best or most easily via (more or less) standardized questioning, some via observation, others via content analysis etc. Often a combination of various methods and procedures will be found suitable.

The choice of methods also depends on the specific purpose (formative/ summative) of the evaluation, though, as well as upon the position and status of the evaluation within the organization (e.g., internal/external) - e.g.:

- Evaluators from outside (external evaluation) may have a stronger tendency towards summative evaluation and "more conservative", that is stricter procedures for proof of effectivity. (Has the programme or the course had any effect? And: What is the strength of this effect ? Or: Comparison of the effects of different programmes/media. What are the (relative) costs ?)
- With internal evaluation, when it is formative, one must make the distinction whether the evaluators are directly involved in the development of the programme or course to be evaluated (i.e. whether they are part of the

10) Cf. Patton, 1987, p. 169: "... evaluators must have a large repertoire of research methods and techniques available to use on a variety of problems. Thus today's evaluator may be called on to use any and all social science research methods, including survey research techniques, social indicators, cost-benefit analysis, standardized tests, experimental or quasi-experimental designs, unobtrusive measures, participant observation, and depth interviewing."

development team) or whether they merely help the development team (by providing information, for example). In the case of a stronger involvement there may be a greater tendency to use qualitative procedures, aimed more at "understanding", and to become *personally* involved - in the sense of an '*action research*' approach - into the development process (cf. the paper by Lorentsen). (On questions of internal and external evaluation and also on the involvement of the evaluators in that which is to be evaluated, see in particular the papers by Rekkedal and Lorentsen; cf. also Rossi & Freeman 1989, pp. 450-452.)

In the choice of the methods and the designs for evaluation studies, too, then, a lot speaks for a pragmatic viewpoint (similar to that in the above mentioned controversy about the 'evaluation paradigm' which is to serve as a basis) and for an orientation towards criteria of how useful and feasible the methods/procedures are; as regards 'designs'[11] for evaluation, Cronbach writes (1982):

> A design is a plan for allocating investigative ressources (Finney, 1956). In evaluation - as in basic science - the designer's task is to produce maximally useful evidence within a specified budget of dollars, a specified number of person-years from the evaluation staff and of person-hours from informants, and other such constraints. 'Maximally useful' is a key phrase. Most writings on design suggest that an investigation is to be judged by its form, and certain forms are held up as universal ideals. In contrast, I would argue that investigations have functions and that a form highly suitable for one investigation would not be appropriate for the next. This is to be, then, a *functional theory of design*."
> Cronbach 1982, pp. 1-2

4 Regarding the conception of this report

Evaluation, respectively evaluation research must not only be geared to that which is to be evaluated (e.g. course or system evaluation), but also to the general aims of the institution as well as to their organizational framework. Aims and organizational framework can, however, differ from institution to institution.

It was, therefore, only natural to view the concepts and the practice of evaluation on an institutional level.

The idea was, then, that members/representatives of different teaching institutions in Western Europe, America, Australia and Asia describe the concepts and methods of evaluation applied by their institution.

---

11) Since this report contains hardly any portrayals of individual inquiries, design questions in the narrower (statistical-technical) sense are only seldom mentioned and covered. - All in all one might say that - at least in the field of the more routine evaluation - descriptive procedures are predominant and experimental, resp. quasi-experimental methods (cf. on the latter, for instance, Cook & Campbell 1979) are generally speaking the exception.

The following 'topics' were suggested to the authors for their papers, but it was left up to them, of course, to take into consideration any other topic which seemed of relevance to them:

- general philosophy of evaluation in their institution
- purpose of evaluation (e.g. the relative importance of formative and summative evaluation)
- evaluation areas/subjects (e.g. the institution as a whole, curricula, single courses/course units, media, student services etc.)
- evaluation methods (e.g. quantitative and/or qualitative approach, action research, social survey, (quasi-) experiment, cost-effectiveness analysis, etc.)
- the outcome variables or criteria of evaluation (e.g. for course evaluation: learners' attitudes to the course, completion rates, costs of course production etc.)
- Evaluation results of more general interest (e.g. attitude to communication technology; group differences: e.g. sex differences; special target groups: e.g. handicapped persons etc.); the main emphasis ought, however, to be laid less on concrete investigation results, but rather on the general concepts and methods of evaluation.

5 Regarding the choice of authors and institutions

Owing to the huge amount and variety of distance study institutions (cf. for instance Doerfert et al 1989 or Rumble & Harry 1982) it was clear right from the start that the selection of institutions and authors for this report could only be very restricted and moreover arbitrary.

The editor wrote at the beginning to representatives of 30 institutions of whom he assumed (or knew) that they worked/had worked in the field of evaluation. All institutions which were to be taken into consideration were supposed to offer programmes/curricula on a tertiary, university level.

It seemed to be of more importance to include representatives of different types of distance study institutions (more or less in the way described in the typology given by Keegan 1990, ch. 8, resp. Keegan & Rumble 1982) than to lay the stress on representativeness in the statistical sense. The selection was supposed

to include a certain diffusion in the geographical positions of the institutions and also to consider countries with differing degrees of industrialization.

Of the papers which are now ultimately on hand, nearly half come from *'autonomous distance teaching universities'*:

- the British Open University (OU-UK)
- the German FernUniversität (FeU)
- the Indira Ghandi National Open University (IGNOU) in India
- the Open University of Israel (O.U.I.)
- the Universidad Nacional Abierta (U.N.A) in Venezuela
- the Universidad Estatal a Distancia (UNED) in Costa Rica.

The remaining papers come from institutions which - to simplify it to a great extent - can be grouped together in the category 'dual mode' institutions[12]:

- the Baptist College in Hong Kong
- the Deakin University in Australia
- the Empire State College (ESC) in the U.S.A.
- the Danish Aalborg University[13]
- the University of Rome, "La Sapienza"
- the Humboldt Universität zu Berlin (HUB) in the ex-GDR
- the Karl-Marx-Universität Leipzig in the ex-GDR
- NKI in Norway (including NKI College of Engineering, NKI College of Computer Science and NKI Distance Education).

The inclusion of two institutions from what used to be the German Democratic Republic (GDR) seemed to be especially interesting because in the ex-GDR (and, similarly, in other ex-"Eastern Block" countries) the type of distance study practised is noticeably different from the type of distance study to be found in western countries, which is mostly characterized by two-way communication conveyed by media. Keegan (1990, for instance) describes the distance study model in the ex-GDR as a "consultation model" [14].

---

12) i.e. they offer, apart from distance education, traditional face-to-face teaching, too.

13) The Aalborg University is one of the constituent partners in Jutland Open University (JOU): JOU is not an independent institution, but was established as a joint venture between, amongst others, the universities of Aarhus and Aalborg. - Cf. the paper by Lorentsen or the EADTU Directory 1988.

14) This is a model with written study materials (conceived centrally for all distance education institutions), which is intended to promote self study; the knowledge gained from this material is deepened in obligatory "consultations" which take place regularly in the form of face-to-face teaching, where the students are given leave of abscence with full pay by the firms/factories which have delegated them to these studies. This leave of absence is granted relatively generously (up to 48 days per year). A mediated form of two-way communication (in the form of written assignments to be submitted, for example) was not

The regional distribution of the papers now on hand has turned out to be rather Eurocentric; unfortunately several institutions in Asia and in North America either failed to react to the invitation or else did not subsequently submit a paper.

6 Regarding the papers

Here to begin with a few general statements about the contributions in this report:

- Several authors stress the importance of evaluation for distance education (like, for instance, Castro or Ganor - and similarly Alavarado et al or Rekkedal - in their papers [15]).

- That which is to be evaluated, and how it is evaluated, is so closely linked to the respective institutional context that the majority of the authors gave a short description of their respective institution and its aims at the beginning of their paper.

- The position and integration of evaluation within the institution or organization is obviously different in the differing institutions: in some institutions (like, for instance, the Academic Quality Control Centre at the UNED in Costa Rica), the evaluation department has a *staff* position and provides advice and assistance to those in *line* positions (the management of the institutions or the person(s) responsible for a programme or for the course coordination); in other institutions the organizational position and the flow of information is less clearly regulated.
  The involvement and active participation of the evaluators in the development of that which is to be evaluated (e.g. a course) varies in a similar way from institution to institution (and sometimes even varies within one institution - depending on the concrete task in hand).

---

intended. The model was extremely successful, judging from the graduation rate of approx. 70%. At the moment distance education in the ex-GDR is undergoing radical change - due, amongst other things, to the fact that the leave-of-absence rules followed up till now (and thus the form of consultations practised so far) have been dropped. Apart from this there is a strong tendency towards more autonomy in each seperate university/college of higher education, so that the central production of learning material for all institutions of higher education will also in all probability be discontinued.

15) Rekkedal: "Evaluation has to be an integral part of and a continuous activity in any educational undertaking." - Similarly also Holmberg (1989, p. 170): "Evaluation is a general educational concern with some special implications for distance education." Cf. further Holmberg (1985, p. 110) or M. Thorpe (1988, ch. 1).

- The openness towards the results of evaluation also seems to vary in the different institutions: Rekkedal, for instance, reports on a very great openness, whilst Bartels laments the lack of interest in the results and mentions tendencies towards reglementation and formalization in evaluation.

- A large amount of topics to be evaluated and methods of evaluation are described.

  What is considered to be the task and the topic of evaluation and how much importance is to be attached to it varies, of course, from institution to institution.

  In most of the institutions evaluation has more than one topic and more than one evaluation level.

  A wide range of areas of application (with similar topics of evaluation: from course evaluation via inquiries on more special questions to system evaluation) are to be found for example in the papers by Bartels, Ganor, Koul, Rekkedal or Woodley.

  References and more or less detailed descriptions of *course evaluation* are to be found in practically all of the papers (in particular detail in that by Bartels or by Woodley). (The older the institutions are, that is the more experience the course developers have, the less importance course evaluation seems to have compared to other tasks of evaluation - cf. for example Bartels or Woodley.)

  '*Student assessment*' is examined in particular by Lehmann & Granger and by Vertecchi.

  References to evaluation of '*support services*', resp. *tutoring/counselling* are to be found in Ganor's paper, amongst others; cf. also Rekkedal or Woodley.

  Carrying out *public relations* work for distance education is also named as a task of evaluation - amongst others by Castro (for Hong Kong) or Koul (for India).

  '*Monitoring*' is covered in quite a few of the papers (in those of Alvarado et al or Woodley, amongst others).

  References to '*impact assessment*' (e.g.: inquiries into the chances of graduates on the labour market) are to be found in the papers by Koul or Woodley, amongst others.

  (An overview of some of the topics mentioned/discussed in the papers is given in Box 7; the list of topics is not complete.)

- In the course evaluation a 'developmental approach' is predominant - often linked to qualitative procedures (e.g. criticism of teaching texts by specialists or didacticians in the respective field and by learners).

- Most of the authors are pragmatic in their choice of methods and approaches (e.g. qualitative or quantitative). That is chosen which seems to be useful and practicable in the respective situation.

Box 7: Overview of topics occuring in the papers
("x" : mentioning / discussion of the topic)

Contribution:	Alva	Bart	Cast	Gano	Koul	Lehm	Lore	Mach	Mohl	Nerl	Rekk	Vert	Wood
topics: peculiarities	1			2	3	4	5				6	7	8
	(numbers: see notes on the following page)												
structure of applicants/ participants[9]		x		x	x	x	x			x	x	x	x
student assessment	x		x	x	x	x		x	x	x	x	x	x
course evaluation[10]	x	x	x	x	x		x	x	x	x	x	x	x
programme / Curriculum evaluation	x				x	x	x		x	x	x		x
Student Services:													
- assignments[11]			x	x	x						x		x
- tutoring/ weekend seminars etc.	x		x	x	x		x				x		x
- Media[12]	x	x		x		x	x				x		x
system evaluation[13]		x	x	x	x		x	x	x	x	x		x
special tasks/ responsibilities:													
- market research for new courses/ programs			x		x								x
- public relations for DE			x		x							x	
- didactic training	x		x	x									
organization[14]	x	x	x			x	x	x	x		x		x
qualitative vs. quantitative methods		x	x				x				x		x
Contribution:	Alva	Bart	Cast	Gano	Koul	Lehm	Lore	Mach	Mohl	Nerl	Rekk	Vert	Wood

continued

Box 7 (*continued*): Notes

1) evaluation as quality control
2) facet approach; evaluation unit has also reponsibilities for didactic training
3) evaluation of IGNOU as well as of other DE-institutions in India
4) Program Effectiveness and Related Costs: PERC-Model; extension of the 'multitrait-multimethod' approach
5) aiming at understanding; qualitative; developmental approach
6) diverse evaluation concepts
7) individual and collective evaluation (assessment)
8) diverse evaluation concepts
9) their background, pre-knowledge; proportion of disadvantaged; etc.
10) formative/summative; continuous; critical commenting; developmental testing etc.
11) correcting, commenting, turn-around time
12) audio-visual; computer; computer conferencing etc.
13) drop out/completion; chances on the labor market; impact assessment; cost-effectiveness etc.
14) internal vs. external evaluation; involvement vs. non-involvement; position of the evaluation unit within the organization

Regarding each paper:[16)]

Alvarado, D'Agostino & Bolaños of the Academic Quality Control Centre of the **UNED** in Costa Rica concern themselves chiefly with course evaluation and emphasize the close connection between 'evaluation' and '*quality control*' (whereby the latter is seen as the narrower term). They differentiate between '*process control*' and '*product control*': whilst the first of these attempts to define 'whether the plans are well-conceived and whether they are implemented accordingly (i.e. whether the tasks are carried out and the implementation of the said plans is performed in the anticipated time and form, and with the desired degree of quality)', the 'product control' intends 'to verify that the intermediate products (mainly didactic material) have the optimal properties to permit the user to acquire the capacities and characteristics established as desirable in the academic plans which support the program'.

The general aim of quality control is the constant improvement of products and processes - if possible, already in the development phase; in this respect quality control and *formative evaluation* are closely linked to each other; quality control also, however, serves to control finished products, and in this capacity is more equal to *summative evaluation*. The authors develop a typology for the tasks of quality control and its purposes (see Table 1 in the paper by Alvarado et al).

To the tasks of the Academic Control Centre belongs - apart from course evaluation - also 'evaluation of learning (i.e. supervising the quality of the

16) The following notes are intended only to give a first impression of the separate papers and to serve as a first orientation-guide. - The differing length of the notes is not at all intended as a valuation.

instruments of learning evaluation)'; the employees of the Centre train course tutors in the elaboration, usage and interpretation of tests.

A frame of reference is also developed which - similarly to Ganor's 'facet approach' (see below) - enables the tasks and formulations of the problems of evaluation to be specified more exactly.

Bartels concerns himself at great detail with questions of *course evaluation* at the German **FernUniversität** and describes the methods used there (a mixture of qualitative and quantitative techniques - like, amongst other things, questionnaires for written inquiries, group interviews, 'course critique', 'mentor reports' - the latter analogous to the 'tutor reports' in Woodley's paper; see below). He describes how methods have changed over the years (towards a stronger stress on the qualitative techniques like 'course criticism'), and how the initial strong interest in the results of course evaluation gradually declined as the course authors gained more routine (similarly Woodley: see below).

Bartels avoids the pair of concepts 'formative/summative' in course evaluation, speaking instead of *'continous evaluation'* (whereby he wishes to indicate - amongst other things - that the process of course evaluation and improvement will really never be completed).

Bartels also deals with aspects of *system evaluation* (drop-out analyses, graduate inquiries or inquiries into study motivation, for instance).

Furthermore, he critically examines the not always unproblematic relationship between evaluators and those who have commissioned the evaluation (the - so to say - 'customers' who are interested in 'buying' evaluation results)[17] and refers to the frustrations which occur if the results of all efforts at evaluation are disregarded, and no attempt whatsoever is made to put them into use to correct what was evaluated.

Angela **Castro**, who recently changed over from the Australian **Deakin University** to the **Baptist College** in Hong Kong, describes the practice of evaluation at Deakin University, as well as the tasks of evaluation for distance study in Hong Kong, for which she presents a plan for *system evaluation* (with reference to Woodley & Kirkwood 1987). This plan includes amongst other things the following steps or components: basic measures of activity, measurement of efficiency, measures of outcomes for students, specifying the programme aims,

17) It is a tradition at German universities for the university teaching staff ("Professoren") to be autonomous in their teaching; they cannot, then, be made to take evaluation results into consideration, for instance, when revising courses.

evaluating policy. Apart from this, suggestions for *formative and summative course evaluation* are made. With reference to the situation in Hong Kong, she further points out that it is also one of the not unimportant tasks of evaluation 'to improve and establish the credibility of distance education'; since distance education in Hong Kong meets with only little recognition, evaluation here has a *public relations function*, too, - a function or duty to establish the credibility of distance education with outsiders. (Koul cites similar duties of evaluation for the Indira Ghandi National Open University, too - see below.)

Margalit **Ganor** of the **Open University of Israel (O.U.I.)** describes the role of evaluation at the O.U.I. as follows: 'The evaluation department has assumed a mediatory role, conceiving its main task as supplying each of the system components with direct feedback that might assist in decision-making and in improving verbal and written teaching processes'. The field of duties of evaluation at the O.U.I. is very widespread and includes, amongst other things, *assessment*, *evaluation of study materials* and *evaluation of tutoring*, and also uses a wide spectrum of techniques and methods (amongst other things attitude questionnaires, observation of tutorials, experimental and correlational studies). Another special thing about the practices of the O.U.I. that is worth mentioning is that the department which is responsible for evaluation also has responsibility in the field of the *didactic training* of the staff, which means that the results of the evaluation can be put into direct use for such training.

Ganor - starting out from Guttman's facet approach - develops a methodological approach which allows the formulation of problems in the evaluation of study components in distance education to be systematized in a 'mapping sentence'. (Regarding the facet approach see, amongst others: Canter 1985; cf. as an introduction to the facet theory also Brown 1985 and Borg 1986).

Koul sketches 7 types of evaluation at the **Indira Ghandi National Open University (IGNOU)** in India, of which only the first four are linked to the activities of the IGNOU as a distance teaching university, whereas the remaining three types are related to the function of the IGNOU as an 'agency for coordinating and maintenance of standards in distance education'. The four types linked to the function as a university are: (1) *evaluation for curriculum development* (a kind of *market research*, intending to identify the groups of clientèle and their needs); (2) *evaluation for admission*: chiefly the development of entrance tests for the applicants, who often do not have the formal prerequisites for admission to B.A. or B. Com. programmes at traditional universities; (3) evaluation of student

performance (*student assessment*); (4) *programme/course evaluation* (the *evaluation of tutors* is also to be included in this field).

The IGNOU has also, apart from its function as a university, the role of a central supervising agency for maintaining the standards for distance education in India. For this task, corresponding evaluation procedures have to be developed, which concern the accreditation of programmes/courses (*evaluation for accreditation*), or which are intended to help in the decision as to which institutions are to be given funds for the development of distance education offers (*evaluation for funding purposes*).

In addition, *impact evaluation* as an evaluation type is also described (criteria used here, amongst others: satisfaction of students; their chances on the labour market; their reputation in comparison to that of graduates from other universities etc.). Impact evaluation also serves 'to find ways and means for establishing distance education ... as a major system of mass and diverse education in India'. The last mentioned task corresponds to that which Castro also describes as being necessary with regard to distance education in Hong Kong.

Lehmann & Granger concern themselves intensively with '*student assessment*', and introduce, in their paper on evaluation at the Center for Distance Learning (CDL) at the **Empire State College** (ESC), the PERC approach (PERC: Program Effectiveness and Related Costs), which is closely related to the aims, respectively the 'mission' of the ESC; amongst these aims one must stress in particular the efforts to attain *individualized education* (made possible through individually designed degree plans and learning contracts). The central question of the PERC approach is: 'What kind of students working with what kinds of faculty in what kinds of learning programs change in what ways at what costs?' Essential characteristics of the PERC strategy are: '(1) it focuses on assessing educational outcomes first and then relates costs to those outcomes; (2) it relies upon a multiple perspectives strategy for assessing student learning; (3) it calls for ... a longitudinal design; and (4) emphasizes the practical utility of these results for academic decision makers.'

The assessment, in which the learner him-/herself is involved, has as one of its more important functions that of putting the learner into a position as to be able to relate the knowledge he/she has gained to his/her own study aims.

From a methodological viewpoint the multiple strategy of assessing student learning can be regarded as a consistent further development of the 'multitrait-multimethod approach' (Campbell & Fiske 1959) - or as a tailoring of this approach to fit the task setting of the ESC.

Annette **Lorentsen** of **Aalborg University** in Denmark describes first of all the development of distance education in Denmark. The Danish conception as a 'communicative distance education concept' is compared with the 'knowledge distributing distance teaching concept' at such institutions as the British OU or the German FernUniversität.

After the presentation of three different 'cases' of evaluation Lorentsen sketches out a 'general philosophy of evaluation' for Danish distance education. This philosophy is characterized amongst other things by process orientation, flexibility, involving the evaluators in the processes to be evaluated and also by usage of qualitative data. With reference to Patton (1980) and Holt (1985) she links up with the 'tradition of anthropological field studies' as well as with 'action research'. The absolute neutrality and objectivity of the evaluators is considered to be an illusion and the advantages of active participation/involvement in the processes to be evaluated are accentuated. In other words: the 'hypothetico-deductive natural science paradigm' is abandoned in favour of a 'holistic-inductive anthropological paradigm'. Referring to Thorpe (1988) a 'developmental approach' for evaluation is demanded, aiming at a continuous development and improvement of teaching and learning.

Arlette Urgelles De **Machado** & Leopoldo **Machado** develop a plan for system evaluation at the **Universidad Nacional Abierta (U.N.A)** in Venezuela, naming as a main field of evaluation at the U.N.A instruction and administration; for both areas evaluation committees have been formed at the U.N.A to coordinate the corresponding activities.

The 'institutional evaluation' at the U.N.A is described by the authors as follows: 'It is an institutional self-study performed by staff; the purpose of such study is to evaluate learning, attitudes, attrition, cost-effectiveness, staff-effectiveness and course design; the key elements are: committee work, standards set by staff discussion; professionalism.'

Möhle & Gerhardt (Karl-Marx Universität Leipzig) and **Nerlich (Humboldt Universität zu Berlin)** describe in their papers the development of distance education in the ex-GDR, and report about the practices used so far in evaluating university distance education.

Möhle & Gerhardt and Nerlich deal with *course/curriculum evaluation* and *system evaluation* as well as with *student assessment*.

These two papers are also of interest due to the fact that they show how evaluation is obviously strongly moulded by society and by the political system as a

whole: it becomes clear with Nerlich in particular that evaluation - in a social system with centralised state-controlled planning (especially, here, in the field of education) - also means an evaluation which is directed towards the aims of just this all-encroaching planning or else plans centrally decided by the state. (This was even encouraged by the fact that teaching materials were worked out centrally in each respective subject for all universities offering distance study courses in the ex-GDR.)

The authors (especially Möhle & Gerhardt), however, also examine the prospective future development (after the unification of the two German states and the changes which this will also bring about in institutions of higher education).

For **Rekkedal** at Norway's **NKI** '*quality control* and *quality development*' are the aims of evaluation.

Rekkedal stresses that an important prerequisite of any attempt at improvement of distance education by evaluation is the *openness* of the institution towards that evaluation (and especially of the person responsible for that institution or for the programme being evaluated).

Rekkedal deals in depth with the relative *pros* and *cons* of *internal vs. external evaluation*. Since different forms of internal and external evaluation are practised at NKI, he is in a favourable position for making comparisons. All in all, it would seem, that he prefers internal evaluation, and believes that it should, ideally, be performed by an independent evaluation unit, which is neither directly responsible for the handling/administration of the programme or course to be evaluated nor involved in it in any way; the evaluation unit should, however, cooperate with those engaged in the development of courses or programmes. (Cf. on this point also Lorentsen.)

At NKI the external evaluation rather serves summative purposes, whilst the internal evaluation is mostly formative and aims at the revision and improvement of that which is being evaluated in each case.

Rekkedal lists a series of fields of evaluation - amongst which are the following:

- evaluating students
- curricula and programmes, courses, and media
- tutor's work
- the whole organization (e.g. via analyses on recruitment, drop out, completion).

Different evaluation criteria and different methods of registering these (like, for instance, questionnaires for students and tutors, face-to-face interviews or interviews given by telephone) are used here.

For purpose of formative evaluation (of courses, for example) Rekkedal considers qualitative procedures ('intensive methods, based on depth face-to-face interviews or interviews given by telephone') to be more useful than large-scale surveys, which permit a quantitative statistical analysis.

Alongside the (more or less routine) evaluation of components (e.g. the teaching material), a series of inquiries also took place at NKI on special themes (for instance on telephone tutoring, on the introduction of the 'personal tutor/ counsellor' or the introduction of new media such as computer conferencing).

Furthermore specific evaluation concepts (like, for instance, *'responsive evaluation'* according to Stake 1976) were tried out, using a variety of methods (amongst others, successive questioning, observation procedures, analysis of loggings, problem solving tests).

Vertecchi, **"La Sapienza" University of Rome**, differentiates between 'individual' and 'collective evaluation': whilst the former refers to the individual student, his/her previous knowledge, his/her learning performance and progress, and also (study-related) attitudes, the latter concerns all the students (as a group). With both forms of evaluation one can basically use the same information (data on performance, taken from tests or exams; data from questionnaires or interviews etc.).

Collective evaluation can, amongst other things, also be used to tailor or to modify a course to suit its addressees.

Both these forms of evaluation can be implemented at the beginning of the distance studies, during the studies and at the end (evaluation at initial stages, in intermediate periods and at final stages). For Vertecchi evaluation, then, includes first and foremost the registration of knowledge and learning performance (in the sense of assessment), and also the opinions and attitudes of the students at the beginning of, during and at the end of their studies.

Woodley describes for the British **Open University (OU)** 'the origins, the evolution and the current status of 'institutional research' - i.e. evaluative research that is carried out by and for the institution itself as a means of self-improvement or ... of self-defence.'

He differentiates two main types of institutional evaluation: system evaluation and course evaluation.

Under the heading "*system evaluation*" he includes:

- basic *measures of activity*: e.g. ascertainment of the number of courses or of applicants
- *measures of efficiency*: e.g. throughput of students; the number of graduates or drop outs; cost-effectiveness. (Such evaluation is often linked with inquiries which go further than a mere data description; intending to uncover the causes - for drop out, for instance.)
- *outcomes*: a check up on 'whether adequate learning has taken place'; alongside the registration of performance (via examinations and the assessment system), investigations to compare the level of knowledge of the OU students with that of students from other institutions should be mentioned here. Of relevance is also the inquiry into the acceptance of OU diplomas or degrees by other educational institutions or by employers; furthermore the taking-over of OU materials by other educational institutions
- *programme aims* (investigations how far the programme aims have been achieved): by analysis of the participant structure, for instance, one can check to what extent so far underprivileged groups are represented amongst the participants ('the OU is committed to greater "open-ness" and to an increase in social equity')
- *policy evaluation*: amongst other things
  - formative evaluation in the form of market research (e.g. in order to test the need for new course offers)
  - inquiries to discover the opinion of the students on specific policy options (e.g. reduced tutorial support)
  - inquiries on how widespread the ownership of specific media (like video recorders or PCs) is
  - testing innovations via experiments: e.g. inquiries into the suitability of the OU for school-leavers
- *organizational evaluation*: evaluating the OU 'in terms of its internal arrangements and procedures' (amongst other things with regard to the financial management; or: monitoring of tutors' marking patterns and the turn-around time for assignments).

Regarding *course evaluation*:

This is divided into formative and summative evaluation procedures, whereby the following aims are defined, in conformity with Scriven (1967):

- *formative evaluation*: providing 'information that can be used during the process of developing or preparing materials or learning experiences'. Methods for

doing this are, for instance, 'critical commenting' by colleagues in the same field or by didacticians, and also 'developmental testing' (trying out draft teaching materials with students).

- *summative evaluation*: providing 'information about how well the 'finished' instruction has worked in normal use.' Information sources and methods for this are: feedback from tutors (comparable to Bartels' 'mentor reports') and feedback from students (mostly in the form of written questionnaires: on the 'extent of utilization', for instance, or on their opinion of the course as a whole or of parts of it; on the 'general style of presentation' and on 'specific content issues' like for example on the understanding of 'key concepts').

Similarly to Bartels for the German FernUniversität, Woodley reports that the significance of course evaluation has waned with the increasing 'age' of the institution.

Woodley goes on to depict *cross-sectional studies* on specific innovations or components which are used in a number of courses (such, for instance, as audio-visual media or computer-assisted learning).

Furthermore Woodley describes a typology (going back to West 1989) of different forms of evaluation according to their general aims and predominant values: '*market evaluation*'; '*liberal evaluation*', and '*radical evaluation*'.[18)]

7 Regarding the wording and arrangement of the texts:

Some of the texts were not submitted in English. They were subsequently translated into English by a 'native speaker'. In addition, some of the authors whose mother tongue is not English asked to have their texts checked by a 'native speaker' as regards the language.

All texts submitted by non-native speakers were, then, checked through by a native speaker. The task of translating and checking the English in the texts was

18) These types are circumscribed as follows: *market evaluation*: 'Here evaluation corresponds to the sort of research done by a profit-maximising manufacturing company. Students are seen as customers and research is devoted to maximising their number and their throughput.' *Liberal evaluation*: 'With this type of evaluation the researcher is the students' friend. Information is gathered on why people drop out, what they think is wrong with the courses, etc., so that the system can be made better for them.' (This type is, then, more or less equal to feedback aiming at improving the courses and the studying conditions.) *Radical evaluation*: 'This can take several forms but essentially involves the evaluator in taking a critical stance concerning which people become students, what they are taught, how they are taught and the effects on their lives.' (e.g.: inquiries on recruitment from underprivileged groups or longterm inquiries on the effect of the studies on the further life of the person in question).

taken over by Mrs. Irene Winter. Each translated or checked text was afterwards returned to the author(s) to be re-checked.

As regards the outward form, the texts were left more or less as they were submitted; all that was undertaken was a certain standardization in the printing (for instance the same kind of type-face for the running text; adjustments in the indentation of paragraphs; adjustments in the method of giving quotations etc.)

In certain places small abridgements or editorial alterations needed to be made to some of the papers; this happened only after consultation with the authors of the respective contributions.

8 Acknowledgements

I am indeed greatly obliged to the authors of the papers for their willingness to cooperate and for the great pains to which they have gone.

I would also heartily like to thank Mrs. Irene Winter for having taken the pains to translate and check the English in the texts.

Also I wish to thank Heidi Hasenpusch, who fed most of the texts into the PC.

For their critical study of my paper and their suggestions on it I wish to thank my colleagues Nicola-Maria Bückmann, Klaus Ebbecke, Helmut Fritsch, Anke Schuemer-Kohrs and Gerd Ströhlein.

References

Abramson, T. (1979): Issues and models in vocational education evaluation. In: Abramson, T., Kehr Tittle, C. & Cohen, L. (eds.): *Handbook of vocational educational evaluation*. Beverly Hills: Sage (quoted from Wittmann 1985)

Bååth, J. (1984): Research on completion and discontinuation in distance education. *Epistolodidaktika*, 1&2, 31-43

Borg, I. (1986): Facettentheorie: Prinzipien und Beispiele. *Psychologische Rundschau*, 37, 121-137

Borich, G.D. (ed) (1974): *Evaluating educational programs and products*. Englewood Cliffs, NJ: Educational Technology Publications

Brown, J. (1985): An introduction to the uses of facet theory. In: Canter, D. (ed.): *Facet theory. Approaches to social research*. New York: Springer, pp. 17-57

Campbell, D.T. & Fiske, D.W. (1959): Convergent and discriminant validation by the multitrait-multimethod matrix. *Psychological Bulletin*, 56, 81-105

Cook, T.D. & Campbell, D.T. (eds.) (1979): *Quasi-Experimentation: Design and analysis issues for field settings*. Chicago: Rand McNally

Canter, D. (ed.) (1985): *Facet theory. Approaches to social research*. New York: Springer

Cronbach, L.J. (with the assistance of K. Shapiro) (1982): *Designing evaluations of educational and social programs*. San Francisco: Jossey-Bass

Cronbach, L.J., Ambron, S.R., Dornbusch, S.M., Hess, R.D., Hornik, R.C., Phillips, D.C., Walker, D.F. & Weiner, S.S. (1980): *Toward reform of program evaluation*. San Francisco: Jossey-Bass

Cronbach, L.J. (1983): Course improvement through evaluation. In: Madaus, G.F., Scriven, M. & Stufflebeam, D.L. (eds.): *Evaluation models. Viewpoints on educational and human services evaluation*. Boston/The Hague: Kluwer-Nijhoff, pp. 101-115

Davis, E. (1981): Teachers as curriculum evaluators. London: Allen & Unwin

Doerfert, F., Schuemer, R. & Tomaschewski, C. (eds.); Bückmann, B., Bückmann, N.-M. & See-Bögehold, C. (co-eds.): *Short descriptions of selected distance-education institutions*. Hagen: FernUniversität/ZIFF

EADTU Directory 1988: see Harry & de Vocht (1988)

Finney, D.J. (1956): The statistician and the planning of field experiments. *Journal of the Royal Statistical Society*, 119, 1-27

Franklin, J.L. & Trasher, J.H. (1976): *An introduction to program evaluation*. New York: Wiley

Fritsch, H. (1990): *Drop ins and drop outs*. Paper presented at the International ZIFF Symposion: Distance Education as Communication: The impact of research and the requirements of practitioners. 27-29 Sept. 1990, Hagen

Glass, G.V. & Ellett, F.S.jr (1980): Evaluation research. *Annual Review of Psychology*, 31, 211-228

Guba, E.G. (1969): The failure of educational evaluation. *Educational Technology*, 9, 29-38

Hamilton, D., Jenkins, D., King, Chr., MacDonald, B. & Parlett, M. (eds.) (1977): *Beyond the numbers game - a reader in educational evaluation*. London: Macmillan

Harry, K. & de Vocht, C. (eds.) (1988): European Association of Distance Teaching Universities: Directory 1988. (EADTU Directory 1988). Milton Keynes, U.K.: International Centre for Distance Learning

Holmberg, B. (1985): *Status and trends of distance education*. Second revised edition. Lund: Lector Publishing

Holmberg, B. (1989): *Theory and practice of distance education*. London: Routledge

Holmberg, B. (1990): *A bibliography of writings on distance education*. Hagen: FernUniversität/ZIFF

Holt, D.M. (1985): Theoretical models of evaluation. *Open Campus No. 11*. p. 3ff. Deakin University (quoted from Lorentsen in this report)

House, E.R. (1980): *Evaluating with validity*. Beverly Hills: Sage

Keegan, D. (1986): *The foundations of distance education*. London: Croom Helm

Keegan, D. (1990): *The foundations of distance education*. 2nd revised edition. London: Routledge

Keegan, D. & Rumble, G. (1982): Distance teaching at university level. In: Rumble, G. & Harry, K. (eds.): *The distance teaching universities*. ch. 1, pp. 15-31. London: Croom-Helm

MacDonald, B. (1973): *Educational evaluation of the national development programme in computer assisted learning*. Proposal prepared for consideration of the Programme Committee of the National Programme (quoted from McCormick & James 1989)

McCormick, R. & James, M. (1989): *Curriculum evaluation in schools*. 2nd edition. London: Routledge

Murphy, R. & Torrance, H. (eds.) (1987): *Evaluating education: issues and methods*. London: Harper & Row

Parlett, M. & Hamilton, D. (1977): Evaluation as illumation: A new approach to the study of innovatory programmes. In: Hamilton, D. et al (eds.): *Beyond the numbers game*. London: MacMillan, pp. 6-22

Patton, M.Q. (1980): *Qualitative evaluation methods*. Beverly Hills, CA: Sage

Patton, M.Q. (1987): *How to use qualitative evaluation methods*. Newbury Park: Sage

Patton, M.Q. (1990): *Qualitative evaluation and research methods*. 2nd Edition. Newbury Park: Sage. (Rev. ed. of Patton 1980)

Peters, O. (1988): Anmerkungen zum Studienabbruch. *ZIFF-Papiere 73*. Hagen: FernUniversität / ZIFF

Reichardt, C.S. & Cook, T.D. (1979): Beyond qualitative versus quantitative methods. In: Cook, T.D. & Reichardt, C.S. (eds.): *Qualitative and quantitative methods in evaluation research*. Beverly Hills, CA: Sage

Rossi, P.H. & Freeman, H.E. (1989): *Evaluation. A systematic approach*. 4th edition. Newbury Park: Sage

Rumble, G. & Harry, K. (eds.) (1982): *The distance teaching universities*. London: Croom Helm

Rumble, G. (1983): Economics and cost structures. In: Sewart, D., Keegan, D. & Holmberg, B. (eds.): *Distance education. International perspectives*. London: Croom Helm

Ruthman, L. (1980): *Planning useful evaluations. Evaluability assessment*. Beverly Hills, CA: Sage

Schuemer, R. & Ströhlein, G. (1990): *Theoretische und methodologische Aspekte der empirischen Dropout-Forschung*. Paper presented at the International ZIFF Symposion: Distance Education as Communication: The impact of research and the requirements of practitioners. 27-29 Sept. 1990, Hagen

Scriven, M. (1967): *The methodology of evaluation*. AERA Monograph Series on Curriculum Evaluation. No. 1. Chicago: Rand McNally

Scriven, M. (1977): Goal-free evaluation. In: Hamilton et al (eds.): *Beyond the numbers game*. pp. 130-134. London: Macmillan

Scriven, M. (1980): *The logic of evaluation*. California: Edgepress

Stake, R.E. (1976): *Program evaluation, particularly responsive evaluation*. Occasional papers series, 5. Western Michigan University, College of Education.

Stake, R.E. (1983): Program evaluation, particularly responsive evaluation. In: Madaus, G.F., Scriven, M. & Stufflebeam, D.L. (eds.): *Evaluation models. Viewpoints on educational and human services evaluation*. Boston/The Hague: Kluwer-Nijhoff, pp. 287-310

Stenhouse, L. (1975): *An introduction to curriculum research and development*. London: Heinemann

Stufflebeam, D.L., Foley, W.J., Gephart, W.J., Guba, E.G., Hammond, R.I., Merriman, H.O. & Provus, M.M. (1971): *Educational evaluation and decision making*. Itasca, IL: Peacock

Suchman, E.A. (1967): *Evaluative research: Principle and practice in public service and social action programs*. New York: Russel Sage Foundation

Thompson, M.S. (1975): *Evaluation for decision in social program*. Westmead, U.K.: Saxon House

Thorpe, M. (1988): *Evaluating open and distance learning*. Harlow: Longman

Weiss, C. H. (1972): *Evaluation research*. Methods of assessing program effectiveness. Englewood Cliffs, NJ: Prentice Hall

Weiß, M. (1975): *Effizienz im Bildungswesen*. Weinheim: Beltz

Wentling, T.L. & Lawson, T.E. (1975): *Evaluating occupational education and training programs*. Boston, MA: Allyn & Bacon (quoted from Wittmann 1985)

West, B. (1989): The UDACE/RAS/R04 Report. *Developing access: a case study of the Open University, with special reference to the West Midlands Region*. (mimeo). Milton Keynes: The Open University

Wittman, W.W. (1985): *Evaluationsforschung - Aufgaben, Probleme und Anwendungen*. Berlin: Springer

Woodley, A. & Kirkwood, A. (1987): *Evaluation in distance learning*. Student Research Centre, Institute of Educational Technology. Milton Keynes: The Open University

Wottawa, H. & Thierau, H. (1990): *Lehrbuch Evaluation*. Bern: Huber

Orientation and Conceptualization of the Academic Quality Control Centre

Jesusita Alvarado B., Giuseppa D'Agostino S.
& Guiselle Bolaños M.

Centro de Control de Calidad Académica (CONCAL),
Universidad Estatal a Distancia (UNED),
San José, Costa Rica.

Abstract

In distance education systems, planning and evaluation play a more prominent role than in conventional systems. We deal here with the experience of course evaluation carried out by the Academic Quality Control Centre at the UNED of Costa Rica. Since evaluation is a fluid field with many models and methodologies, we shall expound first the theoretical framework underlying our strategy of course evaluation.
A summary of the UNED's evaluation experience is then presented. An analysis of this experience leads to the need for reformulating the course evaluation system; finally, this reformulation is outlined.

A. Theoretical Framework

The Universidad Estatal a Distancia (UNED) of Costa Rica has an Academic Quality Control Centre (Centro de Control de Calidad Académica: CONCAL), which arose to meet the need of verifying the quality of the products and processes at each phase of its academic educational mission: design, production, delivery and evaluation of distance courses.

1. Concept

The word "Control" has a direct relation to the area of administration in general and constitutes one of the basic stages of the administrative cycle, which are Planning, Execution and *Control* .

Its scope is inquiry and verification as to whether the processes carried out and the products achieved comply with the preestablished design; with the basic purpose of reorienting, whenever necessary, in view of the results obtained, the activities and elements which generate these processes and products.

The concept of *Quality* refers essentially to the range of characteristics which a process or product should have in order to be considered optimal. Beforehand, one establishes the minimum level for such characteristics in order that the process or

product which attains such a level be considered satisfactory, in view of having reached the expected degree of quality.

When applying the concept of *Quality Control* to the area of education, "one principally seeks to identify the perturbative elements - those which provoke deviations with respect to the desirable characteristics - with the aim of improving the treatment (which the educational process receives) while it is being developed. Thus, the control is exercised during and by means of the stage of execution. In this way it ties in with the concept of formative evaluation. The control also attempts the valuation of the finished product (centering its attention especially on the student; although it may also address the published instructional materials); but in this case the correspondence is rather with the concept of summative evaluation." (Alvarado & D'Agostino 1990).

Consequently, quality control may be located within the evaluative model based fundamentally on congruence and on decision making. This insofar as it checks that resources are employed in the rationally established manner to achieve the desired products; it verifies whether these comply with the preestablished design (*congruence*), and it supplies useful information for taking decisions and steps to rectify the process either during its implementation or when the product is finished.

Within this perspective, quality control is an evaluative action, but not every evaluative action is equivalent to quality control. "The scope of the latter is more restricted; it omits that which is relative to the correctness of the objectives attained with the 'educational design' ('it assumes the pertinence of the proposed objectives, their suitability to certain needs implicit in them and their congruence with the context in which they are proposed': Pereira 1985), and it does not assure whether the most suitable option has been selected ('it does not contribute to diminish uncertainty with respect to other alternative treatments which could generate a more suitable performance of the proposed objectives': *ibid.*). Therefore, it does not extend to determining the timeliness, the pertinence or the range of a specific academic program." (Alvarado & D'Agostino 1990).

In a university context, this situation may in some ways limit the very work of quality control, unless it be complemented with an evaluation which allows the university authorities to find out whether the designed plans were the most suitable and, thus, whether the academic title the university confers reflects the social effectiveness of what is learned.

So, to obtain complete feedback to the teaching-learning process, one must combine the results of control with the results of interpretative evaluations, expert judgments, valuations of collateral effects and other results.

In this view of control and evaluation, the nature of the work which CONCAL has performed and is performing may be summarized as follows: *contributing to*

evaluating academic excellence by means of processes of information and control, in the planning, production, administration and direction of the institutional resources of the University, which allow possible errors to be anticipated and possible problems to be corrected - problems which are derived from unanticipated situations or errors of those responsible - and to verify the final results or performance of the plans of study.

2. Requirements for implementation of quality control

In the previously outlined conception of quality control, its success demands certain requirements, among which the following are of importance:

- objectivity;
- presence of quality parameters;
- comparison of assumptions with results;
- use of results.

Objectivity

This is one of the most important precautions. There exists some resistance to the work of control, which incorporates the dangers of causing sanctions and setbacks for those affected, and of being a means of satisfying personal aspirations of those who exercise it, in both cases detracting from the just aims which the work has in principle.

To avoid these nefarious consequences, the greatest objectivity and a constant striving for impartiality must be norms in the execution of control. Only in this way can this practice meet its authentic aims and avoid the resentment and distrust of those who must submit to it.

Presence of quality parameters

This refers to the need to possess specifications with respect to which control may be exercised. These should be established beforehand by those responsible for the plan of study in the various phases of the educational process. In practice, these parameters must be stated in operational terms; therefore, the objectives of the activities proper to each stage of the process must be made operational.

Comparison of assumptions with results

This is the responsibility of the controlling entity. "The task to be undertaken - comparison - demands the use of a varied range of techniques to determine to what extent the original purpose has been achieved. To execute it, it is important

that the source of information be appropiate to obtaining data related to the specifications, and that the information-collecting techniques be suitable, so that the indicators or variables (dependent, independent or concurrent) be properly measured." (Alvarado & D'Agostino 1990).

Use of results

This is the responsibility of the authorities or the implementing agencies, who, on being informed of the deviations between "assumptions" and "realities", and depending on the magnitude of the deviation, can take action in the short, medium or long terms. Furthermore, they should investigate the factors of non-concomitance between program and execution and apply corrective measures.

3. Location of quality control in the institutional organizational structure

The entity which has been assigned the task of control occupies a certain place within the organizational structure. Its location is influenced by the nature of its mission, or by the type of decision which must be taken based on the results supplied by the control.

Without attempting to cover the whole range of users of control, we mention two of chief significance: the policy-making body and the body which administers the educational process. The controlling organ should be located as a consultative group to the former if the decisions relate to expanding, reducing or changing the academic programs; or as a staff group of the latter if what is desired is to verify operational aspects, such as, for instance, the effectiveness of the technology of producing self-sufficient written materials.

4. Types of quality control and their respective purposes

If we assume that higher education, as part of its academic mission, should offer the country a "social product" - insofar as it should prepare trained professionals, with a high labor market acceptance - then any institution at this level, even one which carries out its tasks through distance education, must endeavour to control the quality of its products (both partial and final) and of the processes which generate them.

To comply with this requirement, several types of control may be practiced; the determination of which type to use depends on two criteria (Pallavicini 1989):

1. the activity over which to exercise control; and
2. the time or stage at which to realize control.

In the context of distance education, and limiting ourselves to teaching and learning, "activity" mainly covers processes and products. *Process control* tries to determine whether the plans are well-conceived and whether they are implemented accordingly; that is, whether the tasks carried out in the implementation of the said plans are performed in the anticipated time and form, and with the desired degree of quality. In this type of control, the basic concern is to follow up the process seeking indications of possible distortion-producing factors; this allows opportune adjustments to be made, which are essential in order to obtain the desired product.

Regarding *product control*, the concern is to verify that the intermediate products (mainly didactic materials) have the optimal properties to permit the user to acquire the capacities and characteristics established as desirable in the academic plans which support the program.

In the same way, with this type one is interested in determining whether the final products (what is learned by the users of the program, whether or not this leads to a degree) possess the aforementioned capacities and characteristics.

The other criterion which allows us to identify types of control is the stage at which control is applied. This parameter allows four classes of control to be distinguished. Control of curriculum design tries essentially to anticipate possible incompatibilities among different parts of the study plan. Also, with control of production of educational materials and control of program implementation, one tries to identify successes and failures in order to make the necessary adjustments. Control of acquired learning helps to identify the effectiveness of the global process. The accompanying table summarizes this classification.

Table 1: Types of Quality Control and their Purposes

Criterion	Type of Control	Purpose
Type of Activity	Process	The identification of errors during the implementation of the educational process and possible solutions to these errors.
	Product	Measure achievements during and at the end of the process.
Stage of Application	Curriculum Design	Anticipate possible incompatibilities among elements of the study plan, and between these and elements which support the plan.
	Production of Educational Materials	Identify successes and failures in production of didactic materials.
	Program Implementation	Identify successes and failures in administration and direction of an academic program or one of its courses.
	Performance	Measure and interpret results of learning: identify effectiveness and decline.

B. A Decade's Experience and its Problems

1. Location of CONCAL in the organizational chart of the UNED

At the UNED, the Academic Quality Control Centre (CONCAL) is assigned as a technical authority (staff position), and not as a line of command, to the Vicerrectoría Académica (Academic Vice-Rectorate). The various dependencies involved in the educational process (curriculum design, production of audiovisual programs, production of printed materials, teaching administration), although possessing administrative autonomy, are assigned to the Vicerrectoría Académica. This particular situation implies that the best location of a controlling organ is as a direct dependent of the Vicerrectoría, so that the results of control flow directly

and quickly to the person who is in the final instance responsible for the educational process as a whole and who exercises authority over all stages of the process.

The connection between CONCAL and the Vicerrectoría, in regard to those aspects whose control is required, is a direct line of access, so that this office can orient the work of the Centre toward such products or processes which it deems necessary to control: the Academic Vice-Rector instructs the Centre as to *what* to control, and the Centre, as the technical authority, decides *how* to control it. As a Control Centre, whose function leads to supplying information about the necessity or otherwise of making adjustments in the educational process, it is one of the supports the Vice-Rector possesses to orient this process.

2. Evaluation tasks

Between 1979 and 1989, CONCAL directed its attention and efforts to critical aspects of the educational activity: *the evaluation of learning and the evaluation of instructional packages.*

2.1. Evaluation of learning

In the control of performance, the most important aspects are, on the one hand, the formative and summative evaluation of learning (self-evaluations, partial and final evaluations: examinations, projects, supervised practice, etc.) and, on the other hand, professional competence in performing the careers for which the university prepares. Due to administrative and other difficulties, the functions of the Centre with regard to the latter task have not yet been established.

In regard to the former aspect (evaluation of learning) the Centre's function has consisted in: *"supervising the quality of the instruments of learning evaluation."* To carry out this function, several specific tasks were established for the professionals employed by the Centre, so as to lend support and consulting in the following fields.

a) Training courses.

Tutors who make evaluations, and heads of courses and careers were trained in the elaboration of tests and the analysis and interpretation of their results. This training was offered as two courses, one on elaboration of tests and another on formative evaluation of learning.

b) Development of evaluation instruments.

When a regular course was given, by agreement among the course head, the evaluating tutor and the CONCAL professional, support was offered in such technical aspects as revision of the balance table of a test, relevance of examination questions, test structure, adequacy of questions as to the type of learning which they evaluate, elaboration of answer sheets, etc.

Initially, this task was realized with most of the University's courses. As the number of courses grew - while the number of CONCAL professionals did not - this service was reduced to some 5 or 10% of the courses given per academic period. This situation required the establishment of selection criteria for courses meriting assistance.

c) Analysis of test results.

For each academic period test results were analyzed, on the basis of the learning objectives, the level of difficulty and the index of discrimination of the items. With these results, a report was drawn up and sent to the head of the course, to enable him or her, together with the tutors, to evaluate those results and to take such corrective steps as were necessary to redirect the teaching process. Running parallel, the Centre compiled a bank of test items for the use of the evaluating tutors and course heads.

2.2. Validation of instructional packages

With respect to the evaluation of instructional packages, there was set the function of *determining the level of academic quality (effectiveness and efficiency) of the instructional packages, and proposing improvements.* To comply with this function, the main task was to advise on the definition and development of the instructional packages and their delivery. This task was effected in the stages of planning, production of the package and delivery of teaching, by means of the following activities.

a) Validation of course curriculum description proposals

As part of the team assigned to this work, consisting of a curriculum designer, the course head, the career head, the academic production member and the CONCAL analyst, the job of the CONCAL member consisted in analyzing the instrument and supervising the aspects of congruence between purpose, objectives and proposed content, as well as the aspects of evaluation laid out in the proposal. This work was designed to ensure the congruence between the elements which

enter in the formulation of an academic program or study plan: *internal validity* of the career structure.

Furthermore, it is necessary to verify that there be a close relation among the professional profile, the course offerings (with their respective objectives and contents), the needs which justify the proposed courses, etc. Since the Centre, for several reasons, has not yet been able to comply with the latter task, at present some alternatives to put it into practice are being studied; these are dealt with in Section C of this document.

b) Evaluation of printed materials

This job was undertaken to verify that the materials were suitable for the methodology of distance education, encouraged individual study, were founded on the appropriate curriculum description, could be mastered in the time allotted and were adequate for the requirements of the relevant discipline. This evaluation was realized at different moments through the following steps.

Validation at the production stage.

The production specialist (not the author) together with the consulting team, of which the CONCAL analyst is a member, endorsed the compliance with the production requirements for printed materials. On CONCAL's part mainly the following aspects were verified: clarity, sequence and sufficiency of content, congruence between objectives and contents, congruence among the intentionally didactic elements of the material, objectives and course content, and technical aspects of the self-evaluation exercises.

Preliminary validation.

Before definitive publication of the materials, these were submitted to a small group of students (4 or 6), in the presence of a tutor trained for this task. The elaboration of instruments to collect information about the quality of the materials, as well as the analysis and interpretation of the results of this validation were the responsibility of the CONCAL analyst, who drew up the final report, so that the necessary adjustments could be made to the materials. The high cost of this type of validation led to the establishment of certain preconditions for its practice: materials previously validated by experts, real possibilities of having the changes included, etc.

Validation in situ.

Its purpose was to determine, from all the students enrolled in the course at a particular time, the efficiency and effectiveness of the instructional package, including the printed materials. The aim was to come to produce suitable and useful materials for a given course. To this end, the following information was collected and analysed.

Students' questions about the themes of the package: The students asked questions of their tutors about these themes; a summary was made of the principal problems the students encountered with the material and it was verified, using analyses of coherence and sufficiency, whether these problems did or did not arise from the form and structure of the material.

Opinions of the tutors: The tutors registered their opinions about the quality, clarity, methodology, timeliness and veracity of the information in each theme of the material which composed the instructional package; their opinions were also sought as to the need for improvements (audiovisuals, self-evaluation exercises, etc.).

Examination results: Tests for two academic periods (semesters) were analyzed, together with the feedback from students and tutors. This analysis comprised a revision of test items for technical quality and an interpretation of the results with respect to the performance of the students in a given theme or subtheme. All this information was then subjected to an expert opinion and conclusions were drawn regarding the effectiveness and efficiency of the instruction.

Expert judgement of a CONCAL analyst: The material was subjected to an internal analysis (coherence and sufficiency) and an external analysis (congruence). Comparing this with the other elements of the analysis, conclusions were drawn which allowed detection of possible successes and failures of the material. With this information, a final report was issued which made recommendations. The decision to redesign the material partially or completely, or to eliminate it, was not the responsibility of CONCAL but of other dependencies. Over 10 years, some 50 reports of this type were issued.

c) Evaluation of audiovisual programs

The radio and television programs were subjected to validation sessions. To this effect, carefully designed instruments were available, whose questions tried to determine whether the program attained its proposed objectives and whether these

were didactic. The CONCAL professional tabulated and analyzed the information and drew up the corresponding reports, which were sent to the audiovisual producers.

3. Valuation of the experience

The results of the work realized by CONCAL over this decade did not always produce the desired effect of raising the academic quality of the courses given by the UNED. Among the factors which contributed to this situation, the following may be mentioned.

3.1. Execution of recommendations

The effectiveness of the results of the work of CONCAL depends to a great extent on the decisions taken, based on the said results, by the organ which CONCAL advises, and their execution by the responsible dependencies.

The experience over the years has not exactly reflected the ideal situation. Decisions as to tasks to be performed by CONCAL in specific cases were often taken on the basis of needs stated by members of the various offices, without intervention of the coordinating authority which should establish priorities among these needs. Moreover, the reports on the results of the evaluations were often sent, not to the decision-making levels, but rather directly to the professionals responsible for the quality of what was evaluated.

3.2. Specification of requirements

As already pointed out under quality control requirements (subsection A.2), it is indispensable, to carry out an evaluation by congruence, that the several dependencies involved in the academic mission define clearly the requirements which the processes and products in their charge must meet. This precondition was not always fulfilled. CONCAL effected such evaluations mainly by expert judgement and performance measurement, participating actively as part of the consulting team in the stage prior to delivery of teaching, and consequently became "judge and party."

3.3. Emphasis on products

It is important that control be directed not only to certain critical points of the instructional process, but also that it cover other related aspects. The evaluation effected by CONCAL has emphasized the quality of products (curricular

descriptions, didactic materials, audiovisual programs, study guides, performance tests), either singly or as part of the evaluation of the instructional package of the courses offered by the UNED. The quality of the processes which generate these and other products was not weighed, nor was there an evaluation of the institutional support services such as student welfare, tutorial service, libraries, study programs, etc. In other words, the mission of CONCAL did not contemplate control of execution and control of performance, concerning professional competence in the discharge of tasks for which the University prepares its students.

The overriding necessity that the work of CONCAL reach its optimal level of effectiveness justified taking the opportunity of an internal reorganization of the UNED to reformulate and reinforce its mission. This reformulation was undertaken recently by the professionals of the Centre and the product of part of this effort is what is presented in the following section.

C. Reorientation of the work of the Academic Quality Control Centre

On the basis of the previously outlined experience, a reorientation of CONCAL's work is at present being proposed, so that, even though its location in the administrative structure of the UNED is unchanged and its general purpose differs little from before - raising academic quality through processes of information and control - some axes inherent in its work have been extended and others introduced. Among these axes, the most prominent are the following:

a) amplification of the range of types of control and thereby of the aspects of the educational process which they cover;
b) valuation directed to processes as well as products;
c) diversification of informational sources;
d) communication of results preferentially to people on decision-making levels;
e) evaluation which emphasizes strongly the analysis of data offered by informants.

To comply with the aims of process and product control in the phases of curriculum design, materials production, execution and educational performance, a methodology is proposed which should, in the final instance, permit the translation of the results into appraisals and quantitative and qualitative measures of easy interpretation.

1. Methodology

The process begins with the identification and definition of the elements which delimit the universe of opinions which the implementation of academic quality control presupposes.

These elements, through the application of a logical nomenclature, are grouped in various classes among which certain logical and useful relations may be established.

We now state several examples of questions which may be used in the various control stages, on the basis of which the instruments of control may be assembled. It should be said that these belong to various categories (users, informants, etc.) and their use depends on the objective of the evaluation undertaken at a given moment.

a) To which dependencies will the results of control be sent?

To those which establish policies or those which administer the educational process. These *users* could be:

a-1. the Consejo Universitario (i. e., the governing body);
a-2. the Vicerrectoría Académica;
a-3. directors of departments and heads of offices;
a-4. heads of academic programs;
a-5. heads of courses.

b) From whom will the required information be solicited?

From professionals involved in the preparation and direction of an academic program or a course in particular, from enrolled students and from prestigious specialists, preferably not connected with the UNED. Some of these *informants* could be:

b-1. authors of course materials;
b-2. authors or consulting specialists for the study plan;
b-3. authors of didactic 'librettos' (screenplays) for radio or TV programs;
b-4. students of courses or careers;
b-5. student counsellors;
b-6. library service personnel;
b-7. presential tutors;
b-8. telephone tutors;
b-9. course heads;
b-10. career heads.

c) What should be evaluated?

In the modality of distance education, the *objects of evaluation* could be one of these:

c-1. suitability of didactic materials;
c-2. validity of an academic program or of a particular course;
c-3. effectiveness of institutional support services;
c-4. academic performance of the students.

d) Which will be the environment of the evaluation?

That in which the object of the evaluation is immersed. This could be:

d-1. a course;
d-2. an academic program or career.

e) Which will be the context for the valuative estimation of the object of evaluation's attributes?

That which allows it to be determined whether what the University offers is what it ought to offer. The following will be *elements of contrast:*

e-1. expectations of the students;
e-2. professional requirements;
e-3. career objectives.

Note that these questions refer to a very general scope of control, which implies that for some of them other questions must be put which should lead to greater specificity. We illustrate this by two examples.

If what should be evaluated is the suitability of published didactic materials, assuming that the production team has already determined the characteristics of quality which these must possess, one may formulate the following question:

f) Which aspects of the object of evaluation should be considered?

f-1. content (depth, range, utility, relevance, organization, ...);
f-2. language (presentation: grammatical construction, vocabulary, style; comprehensibility);
f-3. technico-graphic elements (layout, format, typestyle, figures, ...);
f-4. mathemagenic elements (self-evaluation elements, summaries).

To address the context of valuative estimation (Question e), a question such as this may be posed:

g) What level of user satisfaction should the object of evaluation produce?

g-1. highly satisfactory;
g-2. satisfactory;
g-3. unsatisfactory;
g-4. unacceptable.

Similar questions, related to other aspects, should be formulated in view of the aim of the desired evaluation.

Based on the information collected, various matrices may be constructed which result in divers profiles. Taking for reference the questions compiled so far, if one formulates the matrix:

[a-2, b-4, c-1, d-1, f-2, e-1, g-3],

the following verbalized information is conveyed:

"The Academic Vice-Rector is hereby informed that the students consider that the published material of the course General Administration, in relation to its language (vocabulary, style, etc.), in view of their expectations, is unsatisfactory."

The collection of information will be done by valuative scales of the Lickert type, suitable for the respective type of informant, in which are defined the aspects to be evaluated, the norms used for reference and the indicators; the answers should be YES or NO. Answer sheets designed for optical readers will be used. The usual precautions for questionnaires will be taken: e. g., prior instructions to clear up doubts which could distort the results, selection of the best time to apply the instrument, checks that the forms have the most complete information possible, etc.

The analysis of results will be done electronically. A specially designed statistical program will process the information and design the resulting profiles (line graphs with variables scaled uniformly where possible). Since a profile satisfies its true purpose when compared with a parameter (a normative or ideal profile) rather than when averaged, one must compute, for each profile, coefficients of distance (congruence) and structural similarity, in order to be able to determine coincidence between "observed" and desirable situations.

Finally, the profiles and their coefficients will be printed, and based on the results of these comparisons, the professionals of CONCAL will have to interpret them and prepare reports which orient rational and agile decision-making on the part of the responsible organs.

2. Outlook

Rationality in decision-making is the chief benefit which the Academic Quality Control Centre proposes to offer the university authorities. To this end, the model which it will apply in the evaluation of courses and programs, beginning in 1991, seeks to convert the Centre into a true "transmitting mirror" which will reflect, in an agile and timely way, an image to measure possible distortions in the educational process.

However, in this respect, CONCAL will merely convey information: it will not itself make decisions, nor determine if one option is better than others, nor research into the causes of the deviations: it will only detect them. Nor will it be directly responsible for academic quality, since this is intrinsic to the processes and products, and the production team will be responsible for quality (Alvarado & D'Agostino 1990).

It is the desire of CONCAL to develop, in the medium term, a model which will allow the costs of a course or program to be determined. The university authorities would find it very useful to have available, for a given course or career, simultaneous information about its quality and its cost.

References

Alvarado B., J. & D'Agostino S., G. (1984): El Centro de Control de Calidad Académica. *Revista Enlace* **10**, pp. 46-49. Universidad Estatal a Distancia, San José.

Alvarado B., J. & G. D'Agostino S., G. (1990, in press): El Control de Calidad en la Educación a Distancia. *Revista Educación* **9**, Universidad de Costa Rica, San José.

Bolaños M., G. (1986): *Experiencia de CONCAL en validación de materiales* (preprint). Universidad Estatal a Distancia, San José.

Bolaños M., G. & Brenes E., F. (1985): *Informe Final: Evaluación de Unidad Didáctica de Matemática para el Ciclo Diversificado de la Educación Media* (preprint). Universidad Estatal a Distancia, San José.

Casas A., M. (1987): *Universidad sin clases, educación a distancia en América Latina*. Caracas: OEA-UNA-Kapelusz.

Chávez, J. J. (1989): Metodología empleada en la elaboración del modelo para la evaluación de profesores. In: *Seminario-Taller de Metodología Sistémica para la Evaluación Docente* (preprint). Universidad de Costa Rica, San José.

D'Agostino S., G. & Alvarado B., J. (1982): Evaluación sumativa de los aprendizajes. In: *Seminario sobre evaluación de los aprendizajes* (preprint). Universidad Estatal a Distancia, San José.

Galvis P., A. & González C., M. (eds.) (1983): *Manual de Trabajo del Centro de Control de Calidad Académica*. UNED, San José.

Kaufman, B. (1982): *El control de calidad*. Editorial INDEX, Madrid.

Pallavicini C., V. (1989): *Comportamiento Organizacional*. Editorial de la Universidad Estatal a Distancia, San José.

Pereira, F. (1985): Investigación y evaluación educativa: apuntes para una discusión. In: *Informe de Investigaciones* **1** (5), Universidad Nacional Abierta, Caracas.

Rodríguez D., J. L. (1973): *La función de control en la educación*. C.S.I.C., Madrid.

Rodríguez D., J. L. (1984): *Didáctica General*. Cincel, Madrid.

Weiss, C.H. (1975): *Investigación Evaluativa*. Trillas, México.

Evaluation at the FernUniversität

Jörn Bartels

Division of Evaluation
Center for the Development of Distance Education (ZFE)
FernUniversität,
Federal Republic of Germany

Abstract

This article describes methods and tasks in the field of evaluation which are implemented within the West German FernUniversität at the ZFE. There are two main fields of activities: course evaluation and system evaluation. While course evaluation was in the beginning of the FernUniversität the most important field of work, the emphasis lies now in system evaluation.

Evaluation has always been an integral part of the concept of distance education at the FernUniversität. The constitution of the FernUniversität assigns the task of carrying out evaluation research to the *Center for the Development of Distance Education* (ZFE), focussing on the following areas of interest:

- The evaluation of individual courses with the aim of improving the material sent out to students (*course evaluation*);
- the evaluation of the teaching system, research into study motivation, drop-out, degree students, access to computers, learning conditions, gender, participation in exams, etc. (*system evaluation*).

From the beginning, the philosophy behind all the evaluation carried out by the ZFE has been to contribute to a continuing improvement of the institutional conditions under which the students study at the FernUniversität.

During the initial stages of the development of the FernUniversität, the most important focus of interest was *course evaluation*. For many years now, course evaluation has been de-emphasized as the research interests have increasingly developed into another direction, mainly the evaluation of the teaching system, or *system evaluation*. This shift in emphasis is mainly due to the fact that the expansion of the FernUniversität had slowed down quite dramatically and fewer courses were introduced or rewritten since curricula had been completed for various degree courses.

The author of this article would like to note that it deals only with those evaluation procedures which are implemented within the FernUniversität at the ZFE.

Course evaluation

In principle *course evaluation* can be conducted in either of the following ways:

- during the actual process of developing course material;

or

- after the material has been developed, as a *continuous evaluation* of the course material.

Most evaluation concepts have as their main aim helping to develop course material under optimal conditions neglecting the review of the real conditions under which the course material is actually used. These evaluation concepts involve a certain risk, because in principle they suggest that it is possible to have the full range of knowledge about all future conditions and that it is possible to develop perfect course material for use under all conditions. Other researchers do not accept this point of view on the basis that knowledge will never be perfect and some improvements will always be necessary. Also, parts of the conditions will change, necessitating new didactical decisions.

It is worth noting that course material which is developed under lab conditions has to be improved for use in real study conditions. The reason for this is, that investigations into the behavior of human beings under quasi-experimental conditions yield only limited information about real-life conditions. For this reason, it seems absolutely necessary to evaluate course material under real study conditions because only then will its true value become clear.

In practice this kind of evaluation is limited with respect to the following two points: On the one hand, in most cases, the development of course material takes place under considerable time pressure. Experience shows that even the best planned schedules tend to see the time set aside for course evaluation as disposable time. Under these conditions it is practically impossible to do any evaluation. On the other hand, this kind of evaluation forces the authors of course material to develop a special understanding for evaluation and for the evaluators, who are usually not known to them. This can produce a conflict since the course developers are interested first and foremost in finishing their project.

Continuous evaluation, which aims to improve continuously the already implemented course material, does not know these limitations. If the risk of sending inadaquate course material to the students can be limited, continuous evaluation is preferable. The FernUniversität has taken great precautions to limit this risk, mainly by giving precise instructions to the authors, supervising them continually, evaluating pilot studies, releasing the material only with great care and with expertises, statements, and certification. Some of these precautions were safeguarded in the preliminary constitution of the FernUniversität.

The most important advantage of continuous evaluation is that it stands for the general possibility of improving the study material, and that it produces on the side of the authors a climate of willingness to do revision work. This is important if we assume that development projects do not lead to results which are *right* or *wrong* but to more or less acceptable compromises between goals and frame conditions which are dependent on time and space.

At the FernUniversität courses are developed under the full responsibility of individual professors although the draft course material has to be approved by a department council before being printed. In the first years of the FernUniversität some central institutes, like the ZFE and the *Institute for Research into Distance Study* (ZIFF), also had to approve the course material. But this policy changed because the senior academic staff came to feel that they had developed sufficient competence in writing and preparing course material to dispense with outside *censorship*. At present, therefore, they consider the assistance of the ZFE and ZIFF less important than in the foundation phase of the FernUniversität.

While in the foundation phase a large proportion of course authors had to be drawn from other German universities, the number of courses prepared by members of the academic staff of the FernUniversität has now increased. Nevertheless, there will always be a considerable number of courses written by external authors, because in this way the departments can be more flexible in setting up new courses and in accommodating the plurality of theories or viewpoints essential in academic teaching.

At the FernUniversität, in keeping with German university tradition, the professors take the lead either as authors or as coordinators of material provided by external authors. The collaboration between various specialists during course development is informal. There are specialists for visual design and instructional

development offering their assistance. They are not incorporated into the departments, but are grouped together in the ZFE.

In general, most of the course material of the FernUniversität has gained an excellent reputation. Students and university teachers from other universities often enroll with the FernUniversität just to obtain the course material. As professors in our system compete with their colleagues from conventional universities the level of the course material meets high academic standards but may not be sufficiently student orientated. The low graduation rates at the FernUniversität are probably also an indication of this pattern.

The fact that the FernUniversität took up teaching immediately it was set up precluded a prolonged phase of course development before courses became part of the curriculum. Of necessity, therefore, the continuous evaluation approach was adopted, taking into account that research had to be carried out after a course had been completed. An advantage of this is that the FernUniversität conducts research while the students are actually working on the course. The research does not take the form of lab experiments or other controlled forms of investigations. Rather, it is carried out with students working on the course under 'normal' conditions. The results of continuous evaluation thus reflect the actual study situation of real-life students.

In order to cope with the large number of new or revised course units in the starting phase of the FernUniversität, standard procedures of course evaluation were developed which consisted predominantly of two instruments for routine evaluation. One of these was a *standardized questionnaire*; the other was a student assessment of the course material (*course critique*). These two instruments were used separately or in conjunction with each other.

From the beginning the FernUniversität used a *standardized questionnaire* in course evaluation. The primary function of this questionnaire was to supply preliminary data on the strengths and weaknesses of a given course unit and to point out problem areas which needed to be studied in greater detail. The use of a standard form to collect this data provided at least a rough and ready indication of problems and difficulties encountered by students. Often it served as a basis for planning more detailed evaluation research, being supplemented mainly by the second routine instrument, *course critique*, for a closer investigation into the nature of the difficulties shown up by the questionnaire.

As early as 1976 it became clear that the administration of the standardized questionnaire as well as the data processing were increasingly difficult due to rapidly increasing numbers of students and courses. It was consequently decided to make use of the FernUniversität's *Optical Mark Reader* (OMR) and to design a questionnaire that could be read automatically by this machine. In terms of saving time there were two advantages: Firstly, the data could be processed automatically and thus be made available more quickly either as a basis for revision or for the planning of further research. Secondly, an OMR-questionnaire could be administered directly by the computer center with the help of a computer program.

The standardized questionnaire was sent to all students enrolled on a given course or, in the case of courses with high enrolment numbers, to a sample of students drawn at random. Students received questionnaires for only one course at a time. The reason for this was to increase the response rate. The students were asked to fill in a separate form for each course unit and to return it as soon as they had completed work on that unit. As questionnaires were frequently sent back together with assignments, the rate of response tended to vary with the working discipline of students, i.e. students not taking a course for credit or students not completing a course were likely to be among the non-respondents. In the first academic year of the FernUniversität the response rate in course evaluation averaged at 37 %. Some years later the response rate was less than ten per cent.

The standardized questionnaire covered the following aspects of the course material:

- course design and presentation of the written material;
- previous knowledge of the course content;
- comprehension of the text;
- use of learning aids;
- subjective degree of difficulty;
- time spent on working through the course;
- problematic pages/chapters.

In addition, the respondents were asked to fill in their student status.

After being processed by the Optical Mark Reader and stored on a computer tape, the data were analysed with the help of SPSS (*Statistical Package for the Social Sciences*). There was a standard procedure but often additional analyses

were found to be necessary because of special characteristics either in the course itself or in the student population enrolled on that particular course.

Since the beginning of 1980 there was a growing tendency away from evaluation with standardized questionnaires. The reason was simply that the results of these investigations were misunderstood because they were too general. The authors said that they could not use these results for the revision of the course units. But this evaluation method was never intended to give concrete answers to the problems the students had with the study material. From the beginning, the aim had been to receive a rough indication of problems and difficulties relatively quickly so that it would be possible to conduct a follow-up study on *problem courses* or *problem units* of a course. Since 1984 it had become clear that the university departments no longer wished to have the basic information gathered with this standardized questionnaire. For this reason this questionnaire is no longer in use.

The second instrument routinely implemented in course evaluation is the so-called student assessment of courses, the *course critique*. With this instrument it is possible to provide detailed and specific information which can be used directly as a basis for the revision of course material. In their *course critiques*, the students comment on the quality of the course units. They point out mistakes and passages which were too difficult or confusing or else badly presented; they explain where they should have been given extra help such as graphs, tables, and explanations; they let us know whether the learning objectives were realistic or not.

Unlike the standardized questionnaire, course critique is not implemented as a survey tool. If a given course is to be assessed by students, an average of five to eight of the students enrolled on the course are chosen from a group of volunteers. Participation is voluntary, and the participants are given a small remuneration.

Up to 1985 student course critics received an extra set of the course material. They were asked to write notes and comments in that second copy as they were working through each course unit, adding a summary of the unit after completion. Since 1986 the course critique has been standardized and participants are no longer given a second set of the course material. Instead of this, they receive a prepared sheet of paper for every course unit, and on this sheet they write their general and specific comments to the course units. This revised method allows all interested students to give a course critique, although experience shows that the number of students who like to participate in course critique is very low.

The course critics receive an extensive catalogue of instructions referring to different aspects of this material, such as:

- text comprehension;
- use of learning aids;
- presentation of the contents;
- assignments;
- learning objectives.

In the earlier procedure, a reference copy was prepared for the author after the students had returned their copies with the comments and notes. This guaranteed the anonymity of the participants and provided a handy overview of all the criticism. An author looking through such a reference copy could quickly and easily locate those parts of the text that the students judged to be in need of revision. In the case of the method of course critique which is currently being used, the course evaluation coordinator collates all the information which has been written on the prepared sheets by the participants, and passes it on to the author of the course in question.

Course critique does not claim to be in any way representative, but it does point to some specific problems experienced by some students in their work with the material. It identifies factual errors and mistakes in the manuscript as well as presenting a more subjective picture of the contents and presentation of the material covered. Together, these pieces of information have proved useful for revision.

Since some students felt more comfortable dictating their comments onto a tape rather than writing them down, a pilot project on *oral course critique* was conducted for a period of three years. Participants received blank audio cassettes with some instructions and personalized statements spoken by the evaluator, and then taped their comments as they worked through the course unit. The tapes were then transcribed and a copy of this transcription was given to the author together with other evaluation results.

Most course authors were very interested in this type of course evaluation because some of the students made a very detailed and lively critique. In most cases the *oral course critique* was more fruitful than the written one. But the evaluators had to invest much more time into the production of the oral course critique. As the responsibility for course evaluation changed to other staff

members this very successful pilot project was dropped and a more standardized way of course critique was implemented.

In the past the FernUniversität also made use of *course-specific questionnaires* because authors had asked to have certain aspects of their courses studied in greater detail. *Course-specific* or *unit-related questionnaires* may be used individually or in combination with other research instruments. They tend to require more time to be designed, administered and processed than standard forms.

In the last ten years there has been a trend away from quantitative course evaluation to more qualitative course evaluation. The reason lies in the fact that course authors are more inclined to accept results which can be directly translated into the revision of the course. They tend to be more interested in students' comments and detailed criticism than they are in a generalized statistical presentation of survey findings which do not supply a tangible guideline for concrete changes.

In course evaluation *interviews* are used mainly to gain further information on specific aspects of the course material or as a follow-up to course critique. Usually, they are conducted either as *group interviews* or as *telephone interviews*.

From the beginning, *group interviews* have proved an effective instrument of course evaluation. In contrast to individual interviews, group interviewing is less time-consuming und less costly in terms of money and personnel, while at the same time yielding generally more information in a given session. Especially in cases where students have differences of opinion, group interviewing may produce valuable insights as participants tend to introduce unsolicited information into the argument.

In our experience, group interviews have been most effective with students who have completed work on the course or course unit in question, and who have done a course critique while studying the material. These students have already given special consideration to certain questions and are more aware of the goals of the evaluation.

While group interviews cannot be considered *routine instruments* of course evaluation they have been used often enough to have made the development of an interviewing guideline feasible. This guideline refers to the following aspects of course material:

- the presentation and contents of the course or course unit;
- assignments and exams;
- the use of learning aids and literature;
- the time needed to complete work;
- organizational problems.

While the guideline simplifies group interviews, it can be easily adapted to the needs of a particular course.

Our experience shows that group interviews are only productive if the professor responsible for the course is not present. In cases where the professor participated in the group interview the students tended to be more restrained with their comments because they anticipated that criticism might have negative consequences for exams which they had to take later.

Telephone interviews are used when information is needed from students in a short space of time. Contacting students by telephone is a quick and inexpensive way of collecting data, with the advantage that students are not asked to do any additional travelling in order to participate in an interview. Just answering a few questions on the telephone does not seem to be disruptive, and normally students are quite ready to participate in these interviews. In recent years, though, the use of telephone interviews has been curtailed due to considerations on personal data protection.

In general, *telephone interviews* serve the following functions:

- they may be used to collect additional data on the course material, asking questions not included in the routine instruments;
- they may be used to gain more in-depth information on the students' assessment of courses, asking questions on why they wrote down a certain critical comment or what they meant by a given statement made in their course critique.

Beginning in 1977 the ZFE, in cooperation with the department of economics, developed another instrument of course evaluation, the so-called *mentor reports*. The aim of the mentor reports was to use the experiences of the mentors in the study centers. The mentors received a form on which they had to report about general study problems, problems with the course of study, and about any course-specific problems of the students which had been the reason for counselling in a study center. The mentor reports were only used for a short time because some of the mentors did not report about the problems of the students, giving their own

comments to the study material instead. Since the department was not very much interested in the mentors' comments, this project was soon discontinued.

At the end of this section I should like to comment on some recent developments in course evaluation at the FernUniversität. As mentioned earlier, course evaluation has become less important at the FernUniversität. The reasons for this need some explanation: First of all, this development can be seen as a result of the fact that responsibility for course evaluation within the ZFE has been shifted. While in the first years social researchers were responsible for course evaluation, now the subject specialists (instructional developers) do this job. The social researchers had a more general interest in the evaluation of the courses because they saw the chance of helping the students to improve their learning conditions. Course evaluation was their main field of work. Other activities had a lower priority. In contrast, the subject specialists are more involved in other activities, for example in the production of TV-programs. For this reason, course evaluation has only a relatively low priority in their field of work. The result is that, although course evaluation does exist at the FernUniversität, it is given only a minimum of emphasis. Most of the academic departments agree with this policy because they think that they develop very well prepared study material with only a minimal need for revision. From their point of view, an intensive course evaluation is no longer considered necessary.

New course authors tend to be more interested in course evaluation because they may still be unsure of whether their written courses fulfill the standard of being self-instructing learning material. Authors with more experience in writing a course tend to have the feeling that they know everything about writing a good course. They do not, therefore, attach importance to the results of course evaluation.

Another point is also very important if one wishes to understand the relevance of course evaluation at the FernUniversität. Professors at German universities enjoy a constitutionally guaranteed right to the freedom of research and instruction. This means that a professor may accept the results of course evaluation in the revision of a course, but that he or she does not have to act upon any student criticism, and that the institution cannot sanction any failure to do so. The professor is absolutely free in his or her decision.

In the first year of the FernUniversität, a paragraph was included in the contracts of the external authors which said that the authors would receive the full fee for writing a course only after the results of course evaluation were

incorporated into a revised version. If the author did not accept the critique of students then he or she had to give reasons why the criticism seemed to be irrelevant. If the department accepted these reasons the full amount of money was paid, otherwise the administration made a reduction. The FernUniversität practiced this method for one year only, although from the institution's point of view it seemed to be an effective instrument for improving the study material by putting pressure on the authors. But the FernUniversität could only use this sanction for external authors, not for internal ones. For this reason the course evaluators prefer to work together only with those authors who are cooperative and have indicated their willingness to accept student criticism and to incorporate it into their revision of the study material. To do course evaluation just for the sake of doing something about course evaluation is a waste of time and a frustrating exercise in futility.

System evaluation

Investigations which are made into the teaching system and related problem areas, are counted as part of *system evaluation*. As mentioned earlier the activities in the field of system evaluation have for many years dominated the evaluation work at the FernUniversität. The aim of all investigations is to identify problem areas in the teaching system of our institution and to support the rectorate, academic committees or the departments in the process of making decisions via the results of those investigations.

In general, there are two different procedures in *system evaluation* at the FernUniversität:

- In most cases the ZFE takes the initiative for an investigation. The ZFE discusses the topics and goals of the survey with the participating cooperation partner (department or - if more than one department - rectorate; the rectorate decides about the participation of the senate). The ZFE carries out the agreed investigation independently and reports to the participating cooperation partner (department, rectorate / senate) about the results. The rector has the final decision about publishing the results of the investigation in *system evaluation*.
- The survey is not undertaken on the initiative of the ZFE. It is rather the rectorate, the senate with the committees or a department who engages the ZFE to carry out a survey on topics and goals which they define. Nevertheless, in this case, too, the ZFE carries out the investigation independently. The ZFE reports the results of the investigation to the *client* (rectorate, senate/senate

committee, department). The rector or the department concerned has the final decision on publishing the results.

Experience shows that the procedure of *system evaluation* gets more and more formalized the older the institution FernUniversität. While it was no problem at all to carry out independent system evaluation research projects in the foundation phase of the FernUniversität, we have been facing more and more problems in this respect since the institution has become established. Formal procedures before implementing a new project have developed into a habit. This has advantages as well as disadvantages. If the research is carried out within the agreed framework it is more difficult for members of the institution to argue against specific investigations and their results because they have the support of the official cooperation partner.

The results of investigations in the area of system evaluation have not always achieved those levels of acceptance which the evaluators thought necessary. But there is an interesting development to be observed. While in the foundation phase of the FernUniversität the results of evaluation studies had some influence on those people who had responsibility in the decision-making process, in the consolidation phase this influence was less significant. Recently we have again been observing a shift toward renewed demand for investigations in this field. Furthermore, the results are no longer generally negated. Our explanation for the willingness to accept the results of such investigations is that the process of implementing and handling these projects is now more formalized, and also that there is an increasing insight that such investigations produce a lot of valuable information.

In system evaluation we carry out mainly empirical studies and secondary statistical analyses. Qualitative investigations are the exception. In the case of empirical investigations the main research instruments are questionnaires which include mainly closed questions and, if necessary, some open-ended questions. For qualitative investigations we have developed a catalogue of questions to which the students are asked to react.

As in course evaluation, in system evaluation we also use chiefly random samples of students or, in some cases, the total student population that is to be investigated. Whether or not a sample is drawn is a question of the absolute number of students in a specific survey.

Investigations in system evaluation at the FernUniversität show a broad field of research activities. At the beginning of the existence of the FernUniversität, research into the reasons for and the level of *drop-out* was the most important area, since the newly established university had the wish to improve the study conditions and to reduce internally produced drop-out rates. The willingness to revise the implemented system existed and the university changed its system to the advantage of the students on the basis of the first drop-out studies. In the first five years an annual drop-out analysis was carried out. The effect was that everybody was at least interested in reading the report, but that was all. For this reason the annual drop-out research was discontinued. Now, drop-out research is done irregularly when it is seen to be necessary.

Another field of research in the past was the *study motivation* of distant students. Several investigations were carried out over the years in order to learn more about the subjective study goals which motivate students to study at a distance. The last research project was designed as a postal questionnaire sent to six samples of FernUniversität students who represented different patterns of participation and who might be expected to pursue different goals.

As part of evaluating the study system the issue of *gender* in distance education became increasingly important. The starting point for this concern was the striking under-representation of women at the FernUniversität. In a large-scale research project on the situation of women and men in distance education a survey was carried out. One part of this survey concerned the problems students face when trying to combine work and family commitments with a course of study, another part focussed on the relevance of study centers for the students.

One topic of investigation in the past and also currently is the *graduates* of the FernUniversität. The aim of the studies on *graduates* is to gain insight into the experiences of the graduates both during their studies at the FernUniversität and after completion of those studies. How does the public view the degrees of the FernUniversität, what standing does it have on the labor market? Currently we are engaged in a follow-up study of graduates who participated in a survey in 1985. The main focus of this study is the further personal and occupational development of these FernUniversität graduates, who completed their studies between 1979 and 1985.

The *participation of students in exams* has also been a frequent topic of evaluation research at the FernUniversität. Some research was carried out on the

question of the date on which the students sat the exams and on how successful they were.

Another piece of evaluation research was carried out on students who had finished the initial stage of studies successfully but who did not complete a degree at the FernUniversität. How many of these students continued their studies at a conventional university, how successful were they there? What is their reason for having dropped out of the FernUniversität, and why do most of these students not complete their degree?

In the last couple of years several investigations about access to computers were carried out. The aim of these investigations was to obtain valid information from the students on how many of them already have access to computers, how many are willing and able to buy a personal computer for their study purposes, how students use the computer, and which kinds of software are in use. On the basis of these results academic departments will decide if and where the personal computer can be integrated into the system of teaching at the FernUniversität.

References

Bartels, J. (1984): The FernUniversität of the Federal Republic of Germany. In: *Evaluation of higher distance education results*. International congress of open and distance teaching universities. Madrid, Spain: Universidad Nacional de Educacion a Distancia, 501-524.

Bartels, J. (1989): *Graduates of the FernUniversität. An Evaluation of their Studies and their Experiences after Graduation*. Hagen: Center for the Development of Distance Education.

Bartels, J. (1989): *Methods and Experiences with Course Evaluation at the FernUniversität*. San José, Costa Rica: Paper presented at the 3er Encuentro Iberoamericano de Educacion a Distancia.

Bartels, J. & Peters, O. (1986): The German FernUniversität: Its Main Features and Functions. In: Enckevort, G. van, Harry, K., Morin, P. & Schütze, H.G. (eds): *Innovations in Distance Education*. Distance Higher Education and the Adult Learner, 1, Heerlen, The Netherlands, 97-110.

Bartels, J. & Wurster, J. (1979): *Evaluation. Konzepte und Erfahrungen an der Fernuniversität*. Hagen: Zentrum für Fernstudienentwicklung.

Prümmer, C. von (1983): *Course Evaluation at the FernUniversität*. Hagen: Center for the Development of Distance Education.

Prümmer, C. von (1990): Study Motivation of Distance Students. *Research in Distance Education*, Vol. 2, 2, 2-6.

Prümmer, C. von & Rossié, U. (1987): *Gender-Related Pattern in Students' Choice of Major Subject*. Selected Research Findings. Women in Distance Education 2. Hagen: Center for the Development of Distance Education.

Prümmer, C. von & Rossié, U. (1990): *Value of Study Centres and Support Services*. Selected Research Findings. Women in Distance Education 3. Hagen: Center for the Development of Distance Education.

System Evaluation: Some Reflections on Current Tertiary Practices in Australia and Hong Kong

Angela S. Castro[1]
School of Continuing Education
Baptist College, Hong Kong

Self-evaluation for consolidation

Holt reported that, as far back as 1982, the first Vice-Chancellor of Deakin University urges the University "to catch its breadth, sit back and take stock of its achievements, consolidate them, build on them and correct some of the deficiencies in its practices" (Holt 1987, p. 209). Holt's conclusion in 1987 was that Deakin had not formulated any overall strategy at evaluating its own performance even though it was widely regarded as one of the major distance education providers in Australia. Has the picture changed by 1990? The answer is that it has not although some staff are beginning to see the need for such a strategy and that there is enough expertise and work around which such a strategy could be developed and finally formulated.

Since 1987, evaluation at Deakin has continued to proceed mostly "adventitiously and ephemerally" (Kemmis quoted in Holt 1987, p. 209). The efforts made by individual researchers and course teams and one or two members of the administrative staff have not been part of any deliberate and systematic master plan, and seldom received the official support and recognition that they deserved. One example is the very thorough and sound evaluation of an undergraduate unit in education which would have served an excellent exemplary purpose for other courses, attracted the attention of only a few researchers and remains largely unknown to most within the school concerned, not to mention the rest of the University.[2)]

The academic division within Deakin that has made the most effort at systematic evaluation has been, and still is, the School of Management. Student perceptions of the standards of its off campus Master of Business Administration,

1) Angela Castro has worked at Deakin University as an instructor designer for over 12 years. She is currently Head of Academic Programs in the School of Continuing Education, Hong Kong Baptist College.

2) The unit is Classroom Processes. The chairperson of the team that redeveloped the unit, Professor Rob Walker, who is an esteemed evaluator himself, commissioned three researchers to evaluate the unit's development history, its use of integrated multimedia and study groups.

support methods (weekend schools, study groups, telephone tutorials) are sought on a continuous basis, and recommendations for improvements in evaluation reports are duly studied and adopted. It is regrettable, though, that the results of such evaluations and the subsequent actions taken based upon recommendations are not tabled to the University or widely disseminated outside the School although members involved are not discouraged from writing about them for publications in books and journals.[3)]

The points made so far are to suggest that an university should have a clearly articulated evaluation policy by senior management that can be translated into an action plan. Such a plan should ideally include the following:

- A careful delineation and exposition of the university's framework for evaluation;
- A delegation of responsibility to schools and divisions to undertake both routine and special evaluation of courses and the operational aspects of the administrative subsystems;
- The establishment of earmarked funds to assist with such evaluation;
- The establishment of reporting procedure whereby findings and recommendations may receive the best possible chances of being studied and implemented;
- The establishment of procedures and channels for wider and regular dissemination of evaluation findings as well as the subsequent actions taken to all members of the university involved with distance education activities.

In this way, staff not only become more informed and can learn from one another, but common problems may also be identified to allow concerted central actions to be taken. Over time may be a learning community will become embedded "in the psyche of academic and administrative staff throughout the university" (Holt 1987, p. 209).

The benefits of a proper system evaluation

The views of Bates (1974) on the benefits of a proper system evaluation for the British Open University have particular relevance for Deakin which is now one of

3) Some useful reports which have been published about the MBA evaluation are by Vilgoen, Holt and Petzall (1990) on students' entry perceptions and use of study groups.

eight Distance Education Centres designated by the Federal Government to offer distance education in Australia. This Centre status is to be reviewed in 1994 and the University is acutely conscious that its performance will be under close scrutiny.

The benefits stated by Bates are:

a) It will give the University a much clearer sense of purpose and priorities, bringing out and perhaps even overcoming some of the more intractable and not always recognised barriers to development in the University.

b) It will bring a much clearer sense of reality to what the University, in its present situation, can and cannot do effectively.

c) It will have, as a result of its own evaluation scheme, much greater credibility regarding the usefulness and efficiency of the University than any other comparable institution.

d) By building up a set of codified experience, it will make new course production methods more efficient, more predictable and hence more manageable.

e) By forcing the issues of priorities and effectiveness, it should in the long run reduce considerably some of the intolerable work burdens carried by many staff at the present moment (Bates 1974, pp. 1-2).

Since being granted the new status Deakin has spent its energy at soliciting new opportunities at home and abroad. The University's reputation as a centre of distance education expertise has been considerably enhanced by the contribution of some of its more active schools over the last few years. For example, after having established the credibility of its off campus Master of Business Administration, the School has won three handsome contracts to develop distance education programs for two important national professional bodies, the Australian Society of Accountants and the Australian Association of Professional Engineers. However, the extra workload as a result of these new contracts on some administrative divisions of the University, the Student Centre in particular, has been tremendous. It is not an issue that has received a great deal of attention at senior management level, and further expansion and introduction of new collaborative programs could well cause a great deal of stress for staff and the University may eventualy have to address the situation.

Evaluation at Deakin

76

The lack of an evaluation strategy at Deakin is all the more regrettable because there is considerable expertise existing among staff in its schools and divisions. For example, all five full-time academic members of the Institute of Distance Education are conversant with the theory and practice of evaluation and

have initiated a number of worthwhile projects tied to personal research interests. In particular, two researchers have been able to relate their work to the needs of the academic schools (Management and Education). One has embarked on a longitudinal four-year study of the experience of 18 MBA students for his doctoral dissertation; while another has done a considerable amount of work on the perceptions of teachers as distance students.

The method adopted for most school and individual evaluations has been largely qualitative. Postal questionnaire followed by face-to-face or telephone interviews, with the occasional use of journals and diaries, are the norm. It remains to be seen if Deakin will eventually decide to exploit its staff expertise and enthusiasm to help it assess the critical situations concerning study centres, examination centres, tutor quality, student workload, grading standards of assessment and support to disabled students. A small amount of preliminary work in some of these areas was initiated nearly a decade ago by the first Dean of Educational Services but has not been followed up by any individual or division since then.

A different place, a different pace

The above are personal reflections on Deakin after a twelve year involvement with its distance education research and development. The author has recently been seconded to work in a university in Hong Kong to look after academic programming of distance and continuing education. The first impressions already revealed such tremendous qualitative and quantitative differences in the entire gamut of teaching and learning but it may be premature for the author to make any sort of comparative statements at this stage. What is clear, however, is that evaluation will definitely have a role to play in Hong Kong for one reason alone, and that is, to improve and establish the credibility of distance education, an aspect that has often been taken for granted in Australia.

The "brain drain" phenomenon in Hong Kong, brought on by people's anxiety about the imminent return of the British colony to Chinese rule in 1997, has had overwhelming ramifications for distance education programs offered through the continuing education or extra-mural divisions in the six recognised tertiary

institutions funded by the Universities and Polytechnics Grants Committee.[4] This both influences the quality of the courses thus offered and dictates the curriculum and policy developments at an institutional level.

First, it ought to be stated that while distance education may have established its credibility in Australia, the United Kingdom and elsewhere in the world, it is far from being so in Asian countries such as Singapore and Hong Kong. Travelling through those two places in 1990, the author was surprised by the amount of adverse comments heard and read about distance education in general. While a more thorough investigation into this will definitely be necessary to uncover why the public opinion is so hostile, the author's observations so far seemed to suggest that there could be some serious deficiency in current distance education in Hong Kong. The reasons could be that Hong Kong institutions have not developed enough experience in distance education and have not set up a proper infra-structure to handle it even though the majority of them have had a great deal of experience in offering and operating short, part-time courses to the local adult community. Most institutions have had to buy courses produced overseas which were prepared originally for a particular student profile, a mainly Western culture and social environment, and requiring specific form of tutoring and student support which is often not feasible in Hong Kong. As a result, both institutions and their local students are not finding it easy to make these overseas courses work for them and so the slightest whiff of adverse criticism by students could well be publicised out of proportion in the local press, leading to a public misconception that distance education is "inferior" education.

To placate the distrust and anxiety of the public, none of the Hong Kong institutions really offer distance education in its purest form. Some face-to-face tutoring is always provided to help students in coming to grips with independent study. Nevertheless, what can be certain is that most universities will find it increasingly difficult to maintain face-to-face tutorials for two reasons. Firstly, as space is the most precious commodity in this tiny colony of six and half million people, there is a shortage of suitable venues which can be hired for evening and weekend classes at reasonable rental rates. Secondly, the brain drain means that the pool of qualified and experienced tutors in the market is fast dwindling. It is difficult to hire good tutors and to keep them.

4) The six are: The University of Hong Kong, The Chinese University of Hong Kong, Hong Kong Polytechnic, City Polytechnic of Hong Kong, Baptist College (to be renamed Baptist University in 1991), and the University of Science and Technology (only partially built in 1990).

The number of people leaving to settle overseas in Canada, the United States, Australia and the United Kingdom since 1988 has been running to over 58,000 per year. The Government has taken the remedial measure of increasing the quota of on-campus students at the six universities in an effort to produce more graduates to meet local manpover needs. The indirect consequence of this for continuing education and extra-mural departments is further and stiffer competition for qualified teachers.

Despite the current unfavourable general public opinion about distance education courses, there is no shortage of student demands for distance education courses which will give them an overseas qualification. The popular belief is that it will enhance their chances of being accepted for eventual emigration to the country concerend. Universities' realisation of lucrative market demands has often meant that rapid, expedient decisions are made at the expense of careful curriculum and other educational considerations, which are crucial for ensuring the quality of education provided to the local students.

The entry of Baptist College into the world of distance education

Baptist is the smallest of the six universities in Hong Kong, but it was one of the earliest to take up distance education before the current rush began.

The first was a 1985 colloborative agreement with Ohio University at Athens, United States, to offer two external programs through correspondence to local residents of Hong Kong. They are the Associate Degree of Independent Study and Bachelor Degree of General Studies. All credit earned through the external student programme is considered resident credit and applies toward regular Ohio University degrees.

These programs may be pursued in three different ways:

- Self study using a brief study guide and prescribed textbooks with little to no supervision.
- Independent study using specially structured course material posted with individualised supervision which are regularly posted to students.
- Attending classes conducted in Hong Kong by qualified tutors appointed by Ohio University.

In 1988, the School entered an agreement with the University of Strathclyde in Scotland, United Kingdom, to offer a Master of Business Administration degree program through the distance education mode. This collaboration was reinforced in 1990 when a new Master of Science in International Marketing was offered, also by distance learning. Carefully prepared material is supplemented with regular on-site tutorials given by both Baptist tutors and visiting Strathclyde lecturers.

So far both programs are well-received by students and the Strathclyde courses, in particular, has encouragingly low attrition rate.

Evaluation practices at Baptist College

End-of-term evaluation is carried out on all Baptist College distance courses and its 700 certificate and non-award courses taught part-time. This evaluation is dispensed in the form of a simple questionnaire using multiple choice and a three to five point scale. As the number of students averages 17,000 per term, little qualitative evaluation is attempted by the School of Continuing Education for further in-depth study of the more serious findings discovered in the questionnaire survey.

The people who administer the survey are administrative officers in the School. They are generally inexperienced at formulating evaluation questions and even less so at collating and interpreting data but who are nevertheless given the task simply because they are each nominally in charge of co-ordinating certain courses. For some subjects, retired academics from other institutions with subject knowledge and on-campus teaching experience were brought in to review the curriculum and the performance of the teachers. Again, the skill and reliability of such external assessors are of an unknown quality.

One serious omission in all these routine or special evaluations is the non-involvement of the instructors in questionnaire design, data collection, data coding and analyses. This runs counter to the generally accepted premise in Australia and other parts of the world of teacher involvement[5] but is perhaps understandable in

5) Gregory Pastoll, in a 1985 work about evaluation practices in the University of Cape Town, argued strongly "that the ownership of (and responsibility for) an evaluation exercise should reside with the person for running the course" (p. 291). This has also been the premise adopted at Deakin University, particularly in the School of Management.

the Hong Kong context where the relationship between the instructor and the College and between the instructor and the students is largely of a transient nature.

The above are thumb-nail sketches of some first impressions which, this author has been told, reflect the current situation in other Hong Kong tertiary institutions as well. These institutions are all confronted with the common problems of a shortage of qualified staff, the lack of time, the frantic pace of new course development, and an itinerant student and tutor population.

A model of system evaluation

The situation facing Hong Kong distance education is not an enviable one but is one that Hong Kong institutions will have to learn to deal with. A possible way forward may be the phased in implementation of a program of evaluation with the following goals:

* *Begin with certain basic measures of activity*
  Data should be drawn from administrative records to find out how many courses have been produced, how many students there are, how many applicants had to be turned away.

* *Establish a measurement of efficiency*
  Administrative records should be prepared which could answer questions like: How many students successfully complete the courses? What workload do they carry? What is the throughput of students?

* *Establish measures of outcomes for students*
  These may range from merely an analysis of assignment and examination grades to subjective measures of personal, occupational and educational values and benefits. A particular useful survey for Hong Kong students could be the recognition employers give to qualifications earned through the distance mode.

* *Specify program aims*
  There are specifications of what and whom an institution teaches. Such specifications will be extremely useful in guiding the curriculum and instructional design of courses to be developed.

* *Evaluate policy*

  These may be market research on likely demand for particular types of courses, regular survey of students to monitor the financial impact of study, the impact of reduced tutorial support on learning effectiveness, or experiments or pilot schemes before committing the institution to a final decision of major programming.

* *Evaluate the organisation*

  These have as its foci the subsystems of an instituion which handle the distance education activities. The aspects to be evaluated are their internal organisation and procedures. These could include study centre facilites, tutor marking patterns, turn around time of assignments, and counselling for tutors and staff.[6])

Course evaluation and cross-sectional studies

Woodley & Kirkwood (1987) recommend that course evaluation must also proceed in parallel with system evaluation. This may be conceived as being of two kinds:

* *Formative evaluation*

  This includes critical commenting on draft prepared, developmental testing and revision of draft material in the light of formative evaluation. This could be particularly important to Hong Kong institutions before they commit themselves to purchasing material developed overseas.

* *Summative evaluation*

  This is intended to provide information about a course or material in use. The data is based on feedbacks from tutors and students on the extent of material utilisation, overall view of teaching, presentation format of course material and specific contents issue. Revision of existing material in the light of summative evaluation will be the goal of such an exercise.

Woodley and Kirkwood also suggest that it may be desirable to undertake an evaluation study of a particular innovation or component used in a number of courses.

6) The above are suggestions of a framework provided by Woodley and Kirkwood (1987).

> Such studies are likely to be concerned with drawing out generalisations from the use of a particular aspect of teaching, or with establishing the effectiveness of a particular strategy or teaching medium.
>
> (Woodley & Kirkwood 1987, p. 14)

In reality such innovations are more likely to be local solutions and alternatives of student support for courses bought from overseas institution.

Looking ahead

An evaluation strategy modelled on the above is being drawn up for Baptist College. If adopted, the expectation is that the plan has to be implemented gradually because of the lack of financial, human and time resources. The training of existing staff in evaluation skill will be a matter of top priority. This will be a long and costly process and the temptation to commission bilingual evaluation experts from overseas to undertake the task is strong but cannot be justified both on moral or financial grounds. This author feels that it is important for Hong Kong tertiary institutions to have a sense of loyalty and confidence in its staff who should be given new challenges and responsibilities. Work satisfaction can only come about if individuals feel that they are needed and can contribute towards the goal of the College as a whole. While this may sound altruistic, it has its psychological importance and serves some real needs in a place like Hong Kong as everyone gets ready for 1997.

References

Bates, A.W. (1974): *A proposed programme of evaluation*. Institute of Distance Education. The Open University, Milton Keynes. (Unpublished Mimeograph)

Holt, D. (1987): Evaluation practice at Deakin University: the role of the Distance Education Unit and issues for further consideration. *Higher level distance education: perspectives for international cooperation and new developments in technology*. United Nations Organization for Education Science and Culture and Deakin University, pp. 209-224.

Vilgoen, J., Holt, D. & Stanley, P. (1990): The MBA experience: participants' entry level conceptions of management. *Management Education and Development*, vol. 21, part 1, 1990, pp. 1-12.

Holt, D., Stanley, P. & Vilgoen, J. (1990): Unleashing the forces: face-to-face study groups at a distance. *Distance Education*, vol. 11, no. 1, pp. 125-149.

Pastoll, G. (1985): The list format approach to diagnostic course evaluation: a basis for meaningful teaching improvement. *Studies in Higher Education*, vol. 10, no. 3, pp. 289-300.

Woodley, A. & Kirkwood, A. (1987): *Evaluation in distance learning*. Student Research Centre, Institute of Educational Technology, The Open University, Milton Keynes.

Evaluation at a Distance

Margalit Ganor,
Department of Evaluation and Didactic Training
The Open University of Israel (O.U.I.)

Abstract

This paper presents a model of an evaluation system, designed for a distance teaching institute where no direct feedback is obtained on the quality of the educational input and various people are involved in the teaching process.

Three main spheres of work have evolved:

1. Developing an infrastructure -
   a. The content of evaluation was defined within the framework of L. Guttman's "Facet Approach" (Canter 1985) and implemented in "course maps" through "mapping sentence" technique (Shye 1978).
   b. A computerized system was planned for collecting and storing data.
2. Research and ongoing evaluation - manual and computer reports on central topics were devised and regularly disseminated to staff, followed by periodical surveys and research projects.
3. Counseling and guidance - workshops, individual and group counseling were offered to staff with reference to application of evaluation findings to their work.

So far, some of the benefits of such a model to the functioning of a distance teaching institute could be pointed out.

The mediatory role the evaluators assumed, and the diagnostic nature of the feedback supplied to each staff member assisted in decision making and contributed to improving verbal and written teaching processes.

"Facet Approach" emerged as a comprehensive, constructive framework for instruction and evaluation. Specification of course content and its instructional objectives in "course maps" served as a basis for preparing a teaching syllabus, establishing a computerized bank of questions and assessing all course components.

Evaluation at a Distance

Margalit Ganor
Department of Evaluation and Didactic Training
The Open University of Israel (O.U.I.)

The Nature and Purpose of Evaluation

Evaluation nowadays is becoming an integral part of the functioning of many educational systems. It is particularly essential to institutes based on distance teaching where no direct feedback is obtained on the quality of the educational input (Holmberg 1989). This applies to the Open University of Israel (O.U.I.).

Furthermore, while conventional teaching is usually carried out by one teacher, at O.U.I. various people are involved in the process: course-writers, course-coordinators and tutors. This makes the necessity for feedback which is complete and interrelated even more vital.

The nature of evaluation at O.U.I. has evolved from the specific structure of O.U.I. and its definition as a distance teaching system (Ganor 1990). The evaluation department has assumed a mediatory role, conceiving its main task as supplying each of the system components with direct feedback that might assist in decision-making and in improving verbal and written teaching processes. Thus, the evaluators investigate work "in the field," interpret and assess the information received, forward it to the appropriate people and, finally, provide counsel and guidance regarding the process of integrating their findings into the system. Emphasis is placed on the acquisition of data of diagnostic value. Such data should describe the quality of the instructional component under investigation, while simultaneously providing guidelines regarding possible alterations and improvements.

Main Spheres of Work

The work of the evaluation department has evolved within three main spheres:

- the infrastructure for collecting and disseminating data
- research and evaluation
- counseling and guidance

1. Developing the Infrastructure for Collecting and Disseminating Data

Preparing the infrastructure consisted of defining the contents of evaluation and planning a system for collecting storing and disseminating data.

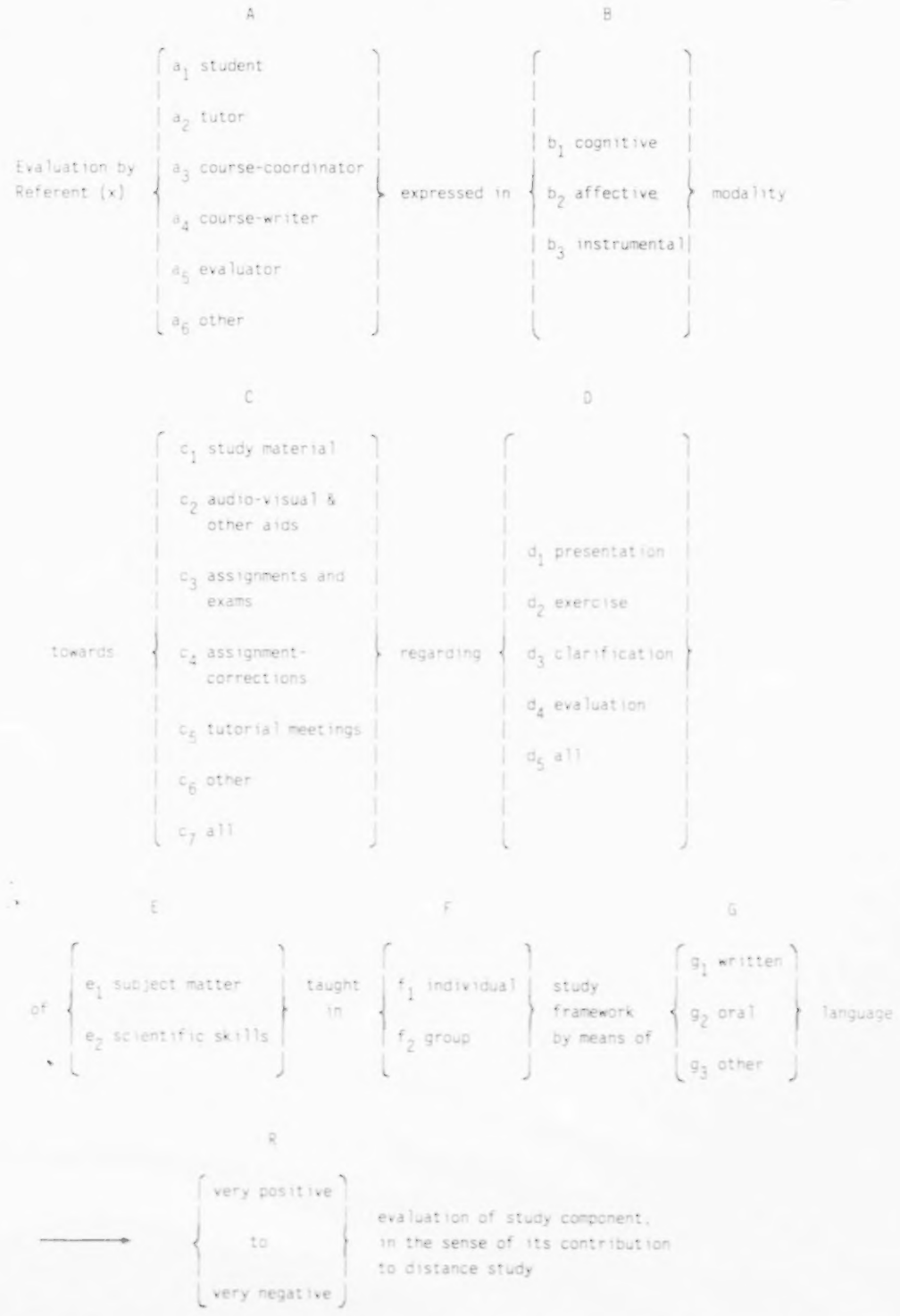
a. Defining the Content

The content of evaluation was defined within the framework of Facet Approach and implemented through Mapping Sentence Technique.

To begin with, a comprehensive definition of goals for evaluation was formed with regard to the nature and function of O.U.I. as a whole. This was followed by a specific definition for the evaluation of the various study components.

The faceted definition presents simultaneously various types of items for evaluation. In it there are eight facets, each of which describes a possible classification of content - study components - and one facet describing the range of evaluation. Each type of items is called a construct and it includes one element of each, or of some, of the suggested facets.

A MAPPING SENTENCE FOR THE EVALUATION OF STUDY COMPONENTS IN DISTANCE EDUCATION



Referent (A) expresses attitudes in modality (B) towards study component (C) regarding didactic aspect (D) of course material (E) taught in study framework (F) by means of language (G). Range of evaluation (R) is from very positive to very negative, in the sense of study component contribution to distance study.

b. Collecting and Storing Data

Data concerning the educational process is currently received from various departments and is stored in O.U.I.'s computer files. Additional information is collected for specific evaluation projects through tools specifically designed for that purpose.

Computer Files

In order to facilitate specific processing of data according to evaluation needs, the team designed three computer files which are reconstituted every term. These comprise three categories of data:

- demographic file - background variables of a current term's students.
- academic background file - a transcript of all courses taken previously by each student.
- grade file - the marks achieved by each student on assignments and examinations in each course of a current term.

Evaluation Tools and Procedures

The distinctive tools and procedures which were devised by the evaluators for collecting data to be used in the assessment process include: attitude-questionnaires, criteria for the evaluation of study units by academic advisors, criteria for observation of tutorials etc. The collection of data from diverse sources is aimed at ensuring an evaluation process which is as objective, unbiased, balanced and validated as possible.

2. Research and Evaluation

While the tasks performed by the Evaluation Department are widely varied, two general directions of emphasis are notable:

a. Current and Ongoing Evaluation

Here the guiding principle for the evaluators is to supply the system with current feedback regarding central topics - in effect "to keep a finger on the pulse." To this end the following manual and computer reports were devised and regularly disseminated to the relevant parties:

- distribution of students in each course according to background variables and progress in studies.
- analysis of C.M.A. (computer marked assignments) results.
- analysis of T.M.A. (teacher marked assignments) results.
- summary report comparing students achievement in C.M.A's, T.M.A.'s and final examinations.

b. Surveys and Research

Research projects undertaken by the department deal mainly with the educational framework of O.U.I.:

- study components - the written text, the role of assignments in distance teaching, etc.
- teaching methods - individual tutoring, group tutorials etc.
- student population - perseverance and achievement of heterogeneous groups, graduates etc.
- teaching staff - tutors in distance teaching - role perception, training, etc.

Projects vary in design and scope. They can be experimental or correlational, some refer to a specific course while others have a wide range with regard to topic, theoretical background, sample, research tools and data time span.

The criteria nature and timing of evaluation are determined in accordance with the objective of the project and in conjunction with the course team.

3. Counseling and Guidance

Considerable importance is attached by the evaluators to didactic aspects of teaching at a distance. It aims to enhance the general awareness of the potential contribution of evaluation to the improvement of teaching. The following activities are undertaken:

- bringing the research findings to the knowledge of the people concerned at O.U.I. and assisting in the application of its findings to the system.
- conducting workshops and offering individual and group counseling (detailed topics hereby exemplified).

The Work of the Evaluation & Didactic Training Department: Examples Relating to Study Components

1. Assignments and Final Examinations

The department deals with the theoretical and technical aspects of composing questions for teaching at a distance (Ganor 1988).

The aims and contents of a given course are re-defined in conjunction with the course team and are presented by Mapping Sentence Technique. This provides a sounder basis for the designing of various types of questions from the cognitive point of view and for ensuring coordination between the course material, the assignments and the exams.

Various "technical" aspects are also taken into account while preparing the testing-system: the optimal number of questions, proportion of optional and obligatory questions, phrasing and multiple choice questions, etc.

Defining distinctly the substance and the cognitive tasks required of the students supplies criteria for assigning weights to the questions, for scoring and correcting them.

Following the theoretical preparatory stage, the department is involved in the more practical part of the process. It offers advice, conducts workshops and monitors achievement. To sum up field activity the evaluators have developed a number of computer reports which would supply feedback on test validity and students achievement.

In view of the large number of questions that must be written each semester, the systematic mapping of the course material as well as of examination particulars and the consequent analysis of the test results have now been utilized to prepare a computerized Bank of Questions.

2. Study Material

Evaluation of study material comprises the following aspects:

- didactic aspects: suitability of material for self-study, degree of difficulty, interest and clarity of texts, graphic presentation and audio-visual aids, etc.
- scientific aspects: a balanced presentation of varied scientific approaches. Updating of material, relevant bibliography and research studies, etc.

Nevertheless, these criteria are to some extent abstract. When evaluating study material it is essential to examine a "live course" in order to assess the effect it has upon students. Hence, data is periodically collected with regard to all course components and is available to course teams.

Demand is also made upon the data for academic reconsideration of courses.

Each course is discussed in the appropriate Sub-Academic committee once every five years to decide whether to continue its teaching or to rewrite it. To enable this Sub-Committee to work efficiently and reach the right decisions the team has established a monitoring-research-evaluation system.

It presents data with regard to the quality of study units and the way in which course objectives have been expressed in "active-teaching" - via assignments, examinations and tutorial meetings.

3. Tutoring

Tutoring is designed to establish a more direct teaching contact with the distant learner. It is carried out through tutorial meetings, individual counseling and correction of assignments. Evaluation of the role of tutoring addresses to the following aspects:

- the content and structure of tutorial meetings
- teaching strategy
- student-attendance at meetings
- individual tutoring
- marking assignments
- experiments with new ways of tutoring

Special evaluation procedures, based upon the research findings, were developed in order to improve tutoring: observation in tutorial meetings, tutor-diaries, videotaping of tutorial meetings, double-checking of final examinations.

The evaluation department conducts workshops attended by course-coordinators and tutors to discuss various aspects of tutoring such as the dynamics of meetings, marking and correcting assignments etc.

References

Canter, D. (Ed.). (1985): *Facet theory. Approaches to social research*. New York: Springer.

Ganor, M. (1988): Assignment construction and evaluation. In: D. Sewart & J.S. Daniel (eds.): *Developing distance education*, pp. 207-210, Oslo, Norway.

Ganor, M. (1990): *On the structure of evaluation at a distance*, 15*th* World Conference ICDE, Caracas, Venezuela. (to be published).

Holmberg, B. (1989): *Theory and practice of distance education*. London: Routledge.

Shye, S. (Ed.). (1978): *Theory construction and data analysis in the behavioral sciences*. San Francisco: Jossey-Bass.

Evaluation at Indira Gandhi National Open University, New Delhi, India

B.N. Koul

Indira Gandhi National Open University (IGNOU)

New Delhi, India

Abstract

The presentation is in three sections. The first of these (A) briefly presents the peculiar responsibilities of IGNOU, which is both a University and an Agency for Coordination and Maintenance of Standards in Distance Education in India. Keeping this dual role in view, the second section (B) presents details regarding evaluation at IGNOU as a University, and then the third section (C) presents briefly the types of evaluation visualized for the second role of IGNOU. Evaluation types detailed in section B are in operation currently and are under constant review for improvements. But, those presented in section C have not gone beyond the conceptual level yet.

A. Introduction

The Indira Gandhi National Open University (IGNOU for short) was established by an Act of the Indian Parliament in September 1985. The first courses of the university were launched in January 1987 and subsequently many more courses have been launched and quite a number are in the pipeline. Being not only a new institution but also an institution with a difference having hardly any models to follow the first five years of the university have experienced a distracting division of attention between the processes of developing appropriate systems for an enormous open educational institution on the one hand and the preparation of courses on the other. Consequent upon this distraction all systems have not developed the way they should have by now. The 'evaluation system' is one of these systems which have yet to develop operationally to a reasonable level of satisfaction. There is yet another point to be taken note of - it is that IGNOU, besides functioning as an open university like any other in the world, has also the responsibility of co-ordinating and maintaining the standards of distance education (DE) in the country, which among other things gives it the powers of disbursing funds to other institutions of distance education in India. With this brief

introductory note on the university, we now reflect on the system of evaluation at IGNOU as it exists today.

At IGNOU the term 'evaluation' is used as a cover expression and very often additional modifiers are used with it to indicate the type of evaluation one may be talking about. The university policy on evaluation recognizes *seven types of evaluation*:

1. Evaluation of social situations and learner needs for purposes of *Curriculum Development*
2. Evaluation of learner competence for purposes of admission to various courses, usually kown as *Entrance Tests*
3. Evaluation of student performance at a particular course for purposes of certification, usually referred to as *Student Assessment*
4. Evaluation of the academic programmes launched by the university, usually referred to as *Programme Evaluation*
5. Evaluation of the *impact* which the IGNOU courses make on the society
6. Evaluation of *courses* prepared by distance education institutions other than IGNOU for purposes of national accreditation
7. Evaluation of *DE institutions* as functional units for purposes of granting funds to them.

The first four types pertain to IGNOU's functions as a university, and the last three types pertain to IGNOU's functions as a national body for co-ordinating and maintaining the standards of DE in India. The first four types are the ones that have been operationalized (though not fully in each and every case), and the last three are yet to be operationalized.

B. Evaluation at IGNOU, the University

In this section we shall touch upon the first four types of evaluation in the order they are given above.

1. Evaluation for Curriculum Development

By evaluation for Curriculum Development we mean those efforts which are directed towards identifying new, unconventional clientele groups for the university. Open and distance education systems make a strong claim for providing socially relevant education. It is common knowledge that today around 60 % of the population of our country suffer from illiteracy and if this illiterate population is to live purposefully, they would need education relevant to their needs, abilities and resources. This education, for obvious reasons, has to be an off-beat type of education. To achieve this objective, the university tries to identify different clientele groups, their needs, the resources available with them for their education etc. and then plans, produces and implements courses for such groups. It needs to be stated that this type of evaluation is needed mainly for professional and awareness courses, while most conventional academic courses are prepared without having recourse to such evaluation.

Once the first stage investigations are made (this is done with the help of questionnaires, study of relevant literature, interviews and field studies) and the generalizations drawn, the preparation of courses is handed over to various Schools of the university. It is obvious that this initial evaluation, which is a kind of market research, leads to the development of new curricula, as such this kind of exercise is seen essentially as an exercise in curriculum development.

2. Evaluation for Admissions

Currently, there are a few programmes offered by the university for which the entry qualifications do not match with what they are at conventional universities. For example, the students who do not have the requisite conventional entrance qualifications for admission to B.A. and B. Com. programmes of the university are allowed to join the programmes provided

- they have attained 20 years of age at the time of admission, and
- they qualify at an entrance test specially designed for the purpose. Normally, it tests a prospective student's general awareness about the environment he/she lives in, his/her language competence and his/her skills at numerals and anylitical abilities.

Success at such tests does not qualify a candidate for any type of certification, instead achieving a predetermined minimum norm at such a test entitles the candidate to admission to the programme for which the test is prepared.

With the passage of time the number and variety of such tests are expected to increase, as more and more students are expected to seek admission to programmes that allow open entry.

3. Evaluation of Student Performance

Generally, at IGNOU, student assessment for purposes of certification may have at least three and in certain cases more than three components. In order to talk about the system with clarity a brief account of course structure and the delivery system in operation at IGNOU is not out of place here.

The master medium used in IGNOU courses is the printed word - self-instructional materials in the form of booklets accompanied by assignments. Usually, less than 20 per cent of study time is to be devoted to audio, video and face-to-face tutorial/counselling sessions. Some courses may have additional components in the form of practical work, field studies, case studies and/or projects.

The first component of student assessment comprises intext questions made available in the self-instructional study materials sent to the students. As a rule answers or suggestions leading to answers to these intext questions are provided alongwith the study material. A student is expected to work on these intext questions on his/her own and having worked on a question look for reinforcement in the answers provided in the materials themselves. These are the questions, responses to which have not to be sent for evaluation to the university or any agency of the university. This component is called '*self-assessment*'. The purpose of this type of assessment is twofold:

- to keep the student on the right track and motivated as he/she works through the study materials, and

- to provide learning activities of various types.

Thus, looked at differently self-assessment in this sense is a type of *Formative Evaluation*.

The second component of student assessment is operated through *assignments*. Students are expected to work on assignments demanding written responses which are to be sent to evaluators working at the study centres of the university. Assignments are usually of two types - Computer Marked and Tutor Marked. (For calculating the overall scores, scores at both types are put together.) These assignments are commented upon and also graded with a view to helping the students improve their performance and also learn about the drawbacks/weaknesses in their presponses. These comments are expected to be elaborate and "teaching in nature". Thus, besides being tools of two-way communication as explained above, assignments serve as tools for students' performance assessment. In other words, assignments are used for two purposes at one and the same time - they serve the purpose of *Formative Evaluation* and also the purpose of *Summative Evaluation* as the grades scored in assignments constitute a component of the overall score a student makes in a course. This component is called '*continuous assessment*'.

The third component of student assessment is based on the term-end-examination which in most cases is a conventional three-hour written examination. This component is through and through "Summative" in purpose.

A course credit at IGNOU is counted in terms of 30 hours of study time, and normally a 4-credit course may be accompanied by anything from 2 to 4 assignments. In the computation of the overall score in a course, the weightage of 'continuous assessment' may be from 25 % to 30 % of the total score and correspondingly, the weightage of the term-end-examination score may be from 75 % to 70 % of the overall score. To get an overall pass a student is expected to get at least a pass grade in both the 'continuous assessment' and the 'term-end-examination'. In most cases the overall pass grade in a course is usually higher than the least pass grade allowed in either component separately.

The University follows a letter grading system of 5 point scale (A, B, C, D & E). The qualitative levels and grade points corresponding to various letter grades are shown in the following table:

Letter	Qualitative Level	Grade Point
A	Excellent	5
B	Very Good	4
C	Good	3
D	Satisfactory	2
E	Unsatisfactory	1

Obtaining at least a D grade in both assignments and terminal examination for each course is essential. However, the integrated average for succeessfully completing a course should be at least C.

Besides the most common system which has been outlined above, various courses may have components that are assessed according to their specific schemes - assessment of practical work in labs, assessment of the projects which the students may have worked on, assessment of case studies which the students may have prepared and other types of group tests given during longer intensive contact programmes.

4. Programme Evaluation

Programme evaluation comprises three components - evaluation of the very *planning of a programme/course*, evaluation of *materials* that result from the plan, and evaluation of the *support services* after the materials/courses have been launched.

(a) Evaluation of the plan for a programme/course:

This activity consists in evaluating a programme/course in terms of the need for it, its marketability, its utility, its economic viability, adequacy and appropriacy of its content and media components.

As an illustration, we elobarate on one of the above factors - the economic viability of a programme/course. For this, various criteria are used to indicate as to how much is to be spent on preparing a course and what might be the returns. In particular attempts are made to see as to how much money needs to be spent on the various operations pertaining to a particular course such as cost of expert-commitee meetings, orientation programmes for course writers, paper and

printing, audio and video materials, support services and the equipment needed for these services, and the cost of learning as far as the student is concerned. These cost-factors help in estimating whether a course is cheaper than the other etc. and thus establish their relative cost-effectiveness. Obviously, such evaluation will yield very useful inputs for planning and decision-making during the 1990s.

From my point of view, we are still at the experimental level, but we are keen to develop this type of evaluation to yield guiding principles to be followed in the future when the resources for course production and implementation will be relatively more difficult to find and they will need to be managed with great care to save the university from possible surprises and difficulties in the future.

(b) Course materials:

To evaluate course materials we need to use the following criteria:

- Is the course content adequate?
- Is the language used appropriate to the level concerned?
- Is the material adequately self-instructional?
- Is the media utilisation pedagogically rational?
- Is the material socially relevant?
- Is the material easily accessible?
- Is the course material considered useful by academics outside the university, and are they using it in their systems?
- Does the material conform to the institutional norms of format, size etc. as required by the university?
- Does the material help the student to get a good pass in the examination?

Programme/course evaluation is a two-tier operation. The first tier consists of those elements of evaluation which are a part of the general management process. For example, the rationale behind and the outline of a particular programme/course is looked into by an expert committee; modifications brought in at this stage are again discussed and reviewed with the course writers; in the process of course writing the materials go through the hands of a

programme/course editor, a language editor, and unit designers, and at the post-production level informal/unsystematic feed-back is obtained from students, counsellors or anybody else. All these stages are the components of programme/course evaluation as a part of the managment process - the establishment has not to put in any separate resources for this level of programme/course evaluation.

However, the second tier of course evaluation may or may not fall within the general management process. The second tier of evaluation may have anyone or more of the following forms:

- Piloting of programmes/courses: In this scheme the materials are tried out with the first batch of students and revisions brought in subsequently. This is a well-known approach and very often the management takes to it almost without giving any thought to its utility in relation to its cost. There are strong reasons why IGNOU should not depend on this approach for each and every course.

  - Given the present constraints on the resources (financial as well as human), it will not be possible for most courses to follow this approach and then offer an improved course to the second batch of students.

  - If the first version of each and every course is prepared hastily and is full of errors, the image of the university will be affected adversely in the long run.

  - It is not advisable, as far as educational ethics is concerned, to use the first batch of students as guinea pigs always and ever.

  - Connected with the above point is the question of the credibility of a course. Should different batches of students get different kinds of courses and yet be assessed by the same examination procedures!

  - Operationally, bringing in revisions immediately after the first launch affects distribution and support services adversely.

  - And lastly, the cost of course design and rate of course production also get adversely affected by this approach.

  It is not suggested that piloting a programme/course is an entirely fruitless task, but it appears that there are better ways of achieving similar results.

- Special evaluation: Special evaluation, whenever needed and proposed, should come in as a project for which the resources are requested for or they are

available from a source other than the university. Obviously, such evaluation will need to have a purpose and, therefore, resources should be available for fulfilling that purpose. Thus, special evaluation of a programme/course needs to be considered on the basis of two factors:

- the evaluation is for a special purpose,
- there are resources available for this special evaluation.

Such evaluation is obiously not a matter of routine. It has to be selective. Some of the cases that may motivate such evaluation are as follows:

- In cases of unusual and inexplicable poor performance of students or that of a programme/course, including adverse reactions by students, academics and counsellors.
- In cases where a new course design is introduced, a radically new target group is catered to or quite an innovative course is launched.
- In cases when a funding authority or an external agency asks for such evaluation with regard to a particular programme/course.
- In case of programmes/courses which involve huge investments (in terms of financial and human resources) and/or take large numbers of students.

- Routine evaluation: Routine evaluation is best done as a part of management process. The advantage of keeping this evaluation within the management process is that in so doing it will not demand high resource allocations. Besides, it will keep the schools in constant touch with the students and the materials, resulting in appropriate and timely feed-back to the schools, course-writers and academic-counsellors. It will eventually feed into the process of course maintenance which can be carried out through supplementary materials without wasting resources. Under routine evaluation
  - each course is evaluated within the year it is launched,
  - in addition to course materials, assignments and the term-end-examination question papers also are evaluated.

If the above two raise issues that cannot be explained easily, special evaluation is conducted.

(c) Student services:

To evaluate student support services we seek answers to the following questions:

- Do these services cater to the information needs promptly, adequately and convincingly?
- Do these services provide for advice at pre-course, on-course and post-course stages adequately and effectively?
- Do these services provide for academic-counselling adequately and conveniently? This includes the effectiveness of tutors with regard to their work on assignment-responses and face-to-face situations.
- Are these services easily accessible?
- Are the staff involved in these services attitudinally tuned to the kind of work assigned to them?
- How do the students rate the support services?
- Do the support services make a pedagogically rational use of various media in operation?
- Do the support services help the learner to get a good pass in the examination?

The criteria that have been listed above bring in a host of variables - human beings in the roles of coordinators, academic-counsellors etc.; geographical situations such as towns, remote areas, rural areas etc.; and pedagogic factors such as classroom techniques, counselling, advising etc. Obviously, evaluation of support services is a difficult area. To make things simpler, this evaluation is considered an aggregate of

- tutor evaluation,
- evaluation of face-to-face sessions,
- evaluation of continuous assessment, and
- the evaluation of support system in general.

Tutor evaluation: Tutors/Academic-counsellors are essentially engaged in four activities:

- they inform, advise and counsel students
- they meet students in face-to-face situations for a number of sessions during an academic year
- they assess assignment-responses
- they help distance learners to learn how they (distance learners) should learn themselves.

These four activities are the bases of evaluating the tutors. This evaluation consists in evaluating

- the nature of communication between them and the students - through letters or in face-to-face situations,
- the quality of their comments on assignment-responses,
- the reliability and validity of the assessment of assignment-responses,
- the turn round rate of assessed assignment-responses.

Evaluation of face-to-face sessions: As far as face-to-face sessions are concerned we evaluate academic-counsellors' punctuality, regularity, nature of rapport with students, quality of the conduct of sessions, students' satisfaction and academic-counsellors' motivation/enthusiasm.

Evaluation of continuous assessment: Evaluation of continuous assessment consists in evaluating the validity and reliability of continuous assessment. This in effect means evaluation of assignments for their validity, their reliability and for their correlation with the questions set in the end-of-the-term examination.

Evaluation of support system: Lastly, the evaluation of the support system itself consists in evaluating the quality and quantity of support available to students in relation to various courses at various study centres. Such evaluation will have implications for recruitment of academic-counsellors, providing support equipment, local and socio-geographical constraints and the corresponding support.

Student support services also constitute a part of the management process. As such no separate resources have to be allocated for such evaluation. Most of this

evaluation should ultimately take place at regional centres; currently, however, it is operated at the headquarters in New Delhi.

C. Evaluation at IGNOU, The Agency for Coordination and Maintenance of standards in DE

As stated earlier IGNOU is a dual institution: (1) a University and (2) an Agency for Coordination and Maintenance of Standards in DE in India.

The last three types of evaluation mentioned in section A (Introduction), belong to the second role of IGNOU. Though a sigificant role, IGNOU has yet to start playing it. Obviously, the three types of evaluation under this head have yet to be operationalized. However, it will not be out of place here to briefly mention all that has been conceptualized so far.

1. Impact Evaluation

By impact evaluation we mean the evaluation that seeks answers to the following types of question:

- Is the student satisfied with what he/she has achieved as a student of the university?
- Can this student sell his/her services on competitive terms in the employment market?
- Is this student rated favourably (if not better) against the students of comparable level from a conventional university?
- Does this student succeed in contributing to his/her social and economic upliftment in particular, and that of the society in general?
- Does this student find easy access to further education in the conventional system, if he/she decides to turn to that system?
- Does this student attract social respect and recognition?

The purpose of such evaluation is to find ways and means for establishing DE firmly as a major system of mass and diverse education in India.

2. Evaluation for Accreditation

One of the ways for maintaining standards in DE is to introduce the system of accreditation which will set up independent bodies to look into the DE courses/programmes launched by various DE institutions in the country. These courses will be evaluated in terms of their objectives, content, nature of study materials, delivery systems, student support systems, and assessment systems on the bases of criteria which have yet to be identified, but may not be different from those used by IGNOU for its own programmes.

The courses/programmes which satisfy an approved minimum standard on this evaluation will be accredited for use throughout the country. The idea is twofold: (1) that a good course/programme produced by a DE institution should be identified and made available for use all over the country, and (2) that 'duplication' should be avoided at all costs - a course/programme that has relevance for the whole country need not be produced at each and every DE institution, instead all such institutions may adopt the relevant course/programme prepared by one of them, provided it is of the required standard. Obviously, evaluation for accreditation has a broader national purpose - maintain standards of DE in the country at the lowest possible costs.

3. Evaluation for Funding Purposes

As a funding agency IGNOU is expected to make funds available to various DE institutions for projects of various types, course development, staff development, estate development and whatever other areas are identified for funding. Though such funding has already started on ad hoc basis, criteria for awards have yet to be identified. Work is on to prepare a workable scheme of evaluation with regard to each of the relevant areas. Once these schemes are ready for use, funding will be done only after each proposal satisfies the norms under the relevant scheme of evaluation.

Assessment at Empire State College -

Strategies and Methods Used in Evaluating Distance Education

Timothy Lehmann
Office of Research and Evaluation (ORE)
Empire State College (ESC)

&

Daniel Granger
Center for Distance Learning (CDL)
Empire State College (ESC)

Introduction: College Mission and Evolving Practice

Ideas about assessment and appropriate strategies, principles, and practices for evaluating student learning have been a part of Empire State College since its founding in 1971. As the College's program was formed in the early 1970's, faculty and administrators debated and experimented with assessment ideas and practices that eventually became part of the culture of the college-assessment of prior learning and assessment of contract learning within the context of individualized education. Since then, the College has refined its mission, elaborated the principles and practices of assessment, and developed more sophisticated and systematic efforts at assessment. This paper describes the assessment strategy, key concepts, methods employed, and findings from illustrative studies involving distance education presently in use at Empire State College.

Any meaningful assessment strategy must start with an understanding of the institutional mission. This is particularly important for an institution like Empire State College, since it has a distinctive mission that combines a number of innovative elements.

Empire State College's central mission is to provide and test effective learning alternatives for postsecondary students. This mission requires:

- *creating alternative structures to increase access for those unable or unwilling to study on campus;*
- *developing academic programs responsive to individual purposes and emerging social needs;*

- *devising teaching and learning methods that serve diverse students with widely varied needs; and*

- *ensuring programs of high quality at reasonable cost.*

In developing the structure, processes, and curricular strategy to fulfill the basic mission, the College holds to three key *educational principles*:

- effective learning derives from the purposes and needs that are important to the individual;

- learning occurs in varied ways and places; and

- styles of learning and of teaching may differ significantly from person to person and from one setting to another
(Empire State College, *The Master Plan 1988-1992*, p. 1).

Over the years, Empire State College has refined its mission and developed a distinctive educational approach which combines the following *innovative elements*:

- individualized education, carried out through a contract mode of learning;

- an open format for access, placing minimal constraints on the time, place, residence, and mode of learning;

- a degree program developed by the student in consultation with faculty, joining the course of study to the student's educational goals;

- a portfolio assessment process certifying prior college level learning;

- a flexible curriculum incorporating eleven broad areas of multi-disciplinary study and modes of inquiry;

- continuing development of learning resources using new pedagogies and technologies (most recently telecommunications); and

- a highly dispersed and decentralized statewide college organization relying for its delivery on a unique mentor-student model (Empire State College, *Self Study Report*, 1979, p. 5 and *Retrospect and Prospect*, 1989, pp. 11-18).

PERC: A Framework for Research and Evaluation[1)]

The Program Effectiveness and Related Costs (PERC) approach of Empire State College, under development since 1972 by the Office of Research and Evaluation, specifically addresses the evaluation issues raised by an individualized educational program for adults. Empire State's program provides for an individually designed degree plan and learning contracts. It also includes procedures for assessing and crediting learning from life and work experiences and makes substantial use of field experience learning. In a fundamental sense, an individualized educational program requires an individualized assessment strategy (Hodgkinson 1975). Until recently however, individualized assessment seemed prohibitively difficult and expensive for most colleges.

The PERC strategy has four features: (1) it focuses on assessing educational outcomes first and then relates costs to those outcomes; (2) it relies upon a multiple perspectives strategy for assessing student learning; (3) it calls for conducting studies within a longitudinal design; and (4) emphasizes the practical utility of these results for academic decision makers (Lehmann 1981, p. 757).

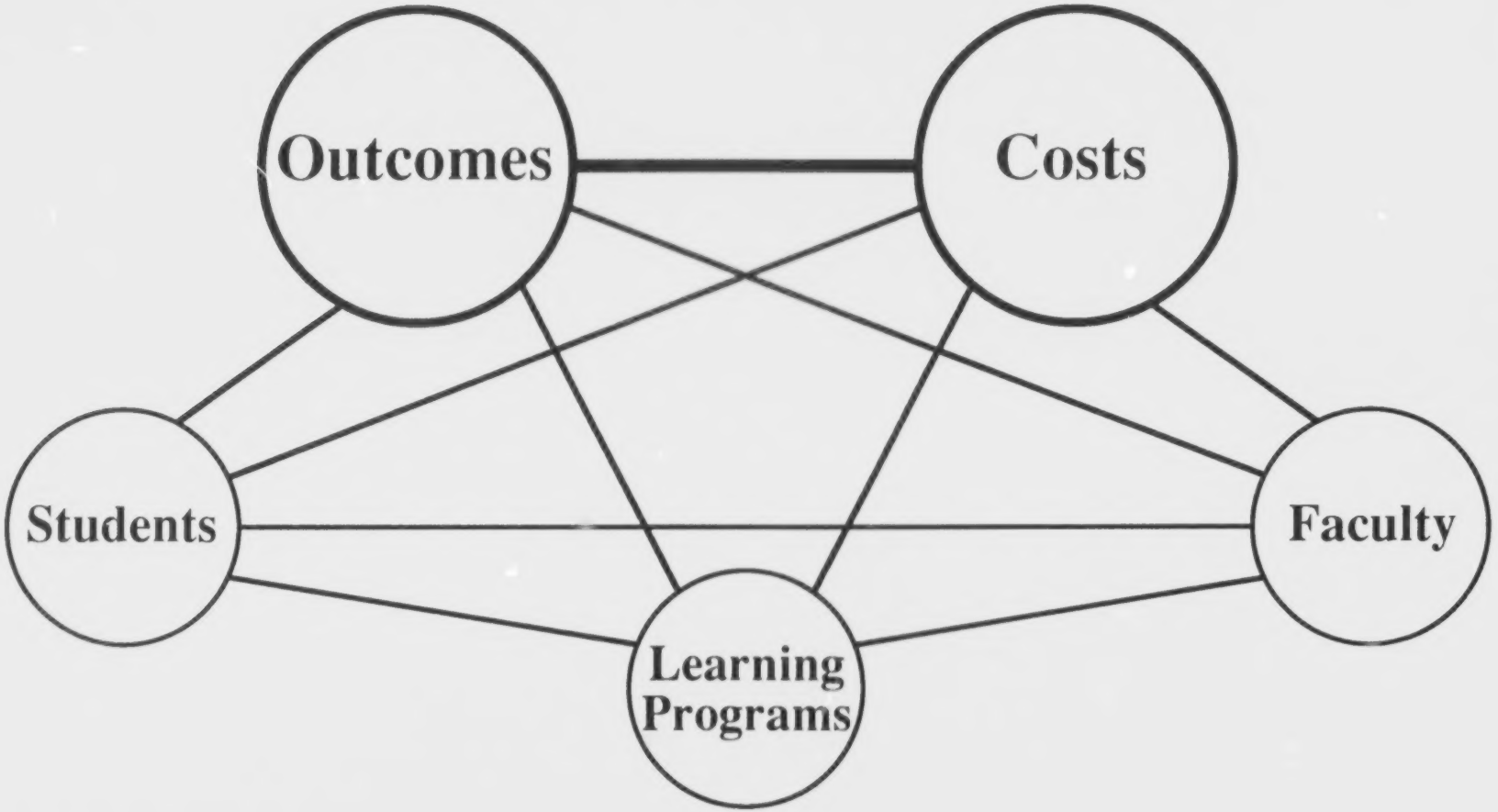
Regarding educational outcomes and costs, the *basic evaluation question* is:

What kinds of students working with what kinds of faculty in what kinds of learning programs change in what ways at what costs?

Embedded in this master question are the five *core variables* of PERC - outcomes, costs, students, faculty, and programs. Among these five variables, the primary relationship is the relationship between outcomes and costs (see Figure 1, next page). In contrast to much of the previous evaluation work that stressed cost-effectiveness or cost-benefit approaches, PERC places primary emphasis upon student learning and educational effectiveness and later introduces the costs associated with a given student's learning or more aggregate measures of the educational program. All too often a focus on costs first leads to a "lowest common denominator" style of decision making, which may shortchange programs that are genuinely effective.

1) This research evaluation model was supported by a major grant from the Fund for the Improvement of Postsecondary Education (FIPSE) and was created under the leadership of Ernst G. Palola who was head of the Office of Research and Evaluation at Empire State College from 1972 to 1980.

Figure 1. THE FIVE VARIABLES OF PERC



Source: Palola and others, 1977.

Lehmann & Granger, Empire State College, USA / 105

The most fundamental assessment activity is to determine what learners are learning (knowledge, skills, values, behaviors) and whether they are achieving their educational goals. The first variable of the PERC framework is outcomes. Each college must identify its own set of outcomes appropriate to the College mission, the kinds of students it attracts, and the kinds of educational programs it offers.

At Empire State, we started by defining outcomes based on learner goals and objectives. Each student designs an individual degree program that provides the basis for assessing learning. Originally, the PERC framework classified student goals into eight *outcomes categories*. These categories were: substantive knowledge, communication skills, cognitive outcomes, developmental outcomes, personal outcomes, occupational outcomes, public service, and unanticipated outcomes (Palola 1977, p. 3).

The most recent formulation recasts the College's goals and outcomes into five broad categories. The primary educational goals of Empire State College are to help students to:

- Enhance abilities as self-directed learners, including setting and stating goals, developing ways to meet them, refining and modifying goals, accomplishing goals and evaluating that accomplishment.

- Develop independent study skills, including the abilities to assess one's learning style and strenghts; locate and use information from various sources; organize study activities; and evaluate study skills.

- Enhance general academic skills, including skills in reading, understanding and interpreting materials of some complexity; writing coherently, correctly, and effectively; speaking and listening; thinking analytically, reasoning with quantitative, mathematical and scientific materials; and using information, inquiry and research skills.

- Gain an understanding of facts, theories, and methods of inquiry and practice appropriate to a field of concentration and to the level of the degree; and develop research, analytical, evaluative, and practical skills appropriate to the field and degree.

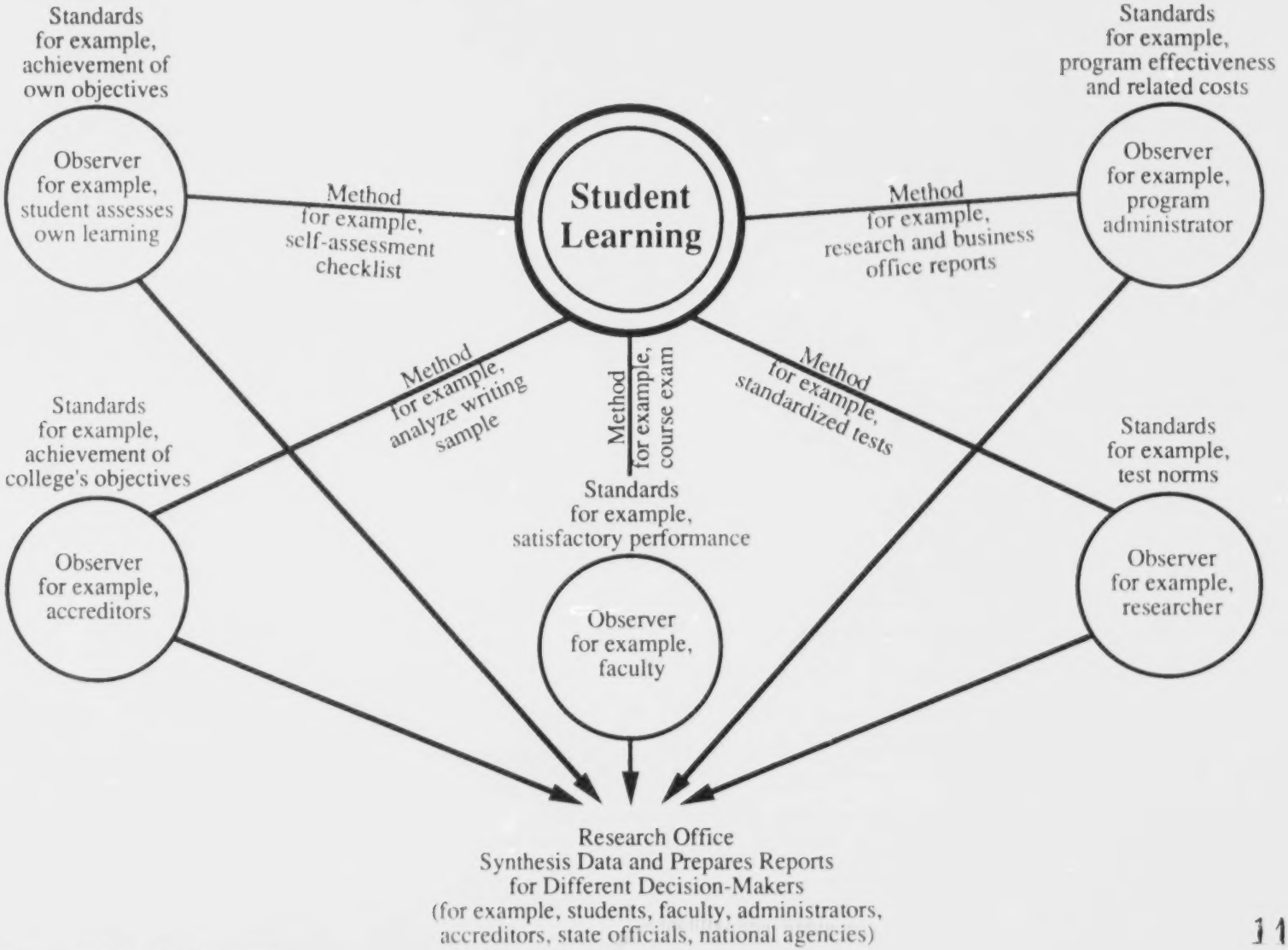
- Broader understanding of a variety of perspectives (e.g., international, gender-based, multicultural, esthetic, historical, literary, scientific, ethnic) that can yield insights about human behavior, society and the natural world and foster the ability to work well with diverse groups.

(Empire State College, Office of Academic Affairs, *Plan for Outcomes Assessment*, June 1990, p. 3).

The inverted Olympic design constituting the PERC framework stresses the outcomes-cost relationship as primary, although all five variables are highly interrelated as shown in Figure 1. In this overview essay, it is not possible to describe in detail the definitions of each PERC variable and how they have been applied over the past 20 years. Those definitions and applications are found in the more than 200 reports and studies conducted since 1972 using the PERC framework. For the interested reader, a list of those studies and reports is found in the Office of Research and Evaluation's Reports and Publications (ESC 1989).

The PERC model addresses the broad range of educational outcomes by means of a *multiple perspectives assessment strategy* featuring five components: (1) multiple observers of student learning; (2) multiple methods of assessing student learning; (3) multiple decision makers using data relevant to policy questions they face; (4) multiple time periods for measuring change in student learning; and (5) multiple standards of assessment (Palola et al. 1977; Lehmann 1981, 1989). This strategy recognizes that student learning is a complex, dynamic, multifaceted, and ongoing process that no single observer, method or instrument can fully capture. Therefore, several observers are required to make assessments of student learning, including the student, using various methods of evaluation. For example, the student is at the center of the assessment process, as shown in Figure 2, next page, because the student is the central participant in the educational process and as such is in especially advantageous position to evaluate certain aspects of his/her own learning in terms of his/her own objectives for learning. Thus, adults may evaluate their own learning on such outcomes as substantive knowledge in a discipline, acquired occupational skills, or enhanced interpersonal competence. Faculty can then evaluate students on the same outcomes and provide another perspective on what they have or have not learned. We expect to find general agreement on most outcome assessments but there are often unanticipated outcomes that emerge via different observers of the teaching-learning process.

Figure 2. PERC'S MULTIPLE PERSPECTIVES STRATEGY


Standards for example, achievement of own objectives
Observer for example, student assesses own learning
Method for example, self-assessment checklist
Student Learning
Method for example, research and business office reports
Standards for example, program effectiveness and related costs
Observer for example, program administrator
Method for example, analyze writing sample
Method for example, course exam
Method for example, standardized tests
Standards for example, achievement of college's objectives
Observer for example, accreditors
Standards for example, satisfactory performance
Observer for example, faculty
Standards for example, test norms
Observer for example, researcher
Research Office
Synthesis Data and Prepares Reports
for Different Decision-Makers
(for example, students, faculty, administrators, accreditors, state officials, national agencies)

Source: Palola and others, 1977.

Other observers of learning outcomes, shown in Figure 2, are the researchers, program administrators and outside evaluators such as accreditors, external examiners, and state education departments (see Association of American Colleges, *Project on External Examiners*, 1987). The research staff using PERC may conduct its own evaluation of student learning using, for example, standardized tests of special exams to measure student mastery of subject matter knowledge. More importantly, the research office has the responsibility to bring together data provided by several observers, analyze the data, and prepare reports that can be used by different decision makers.

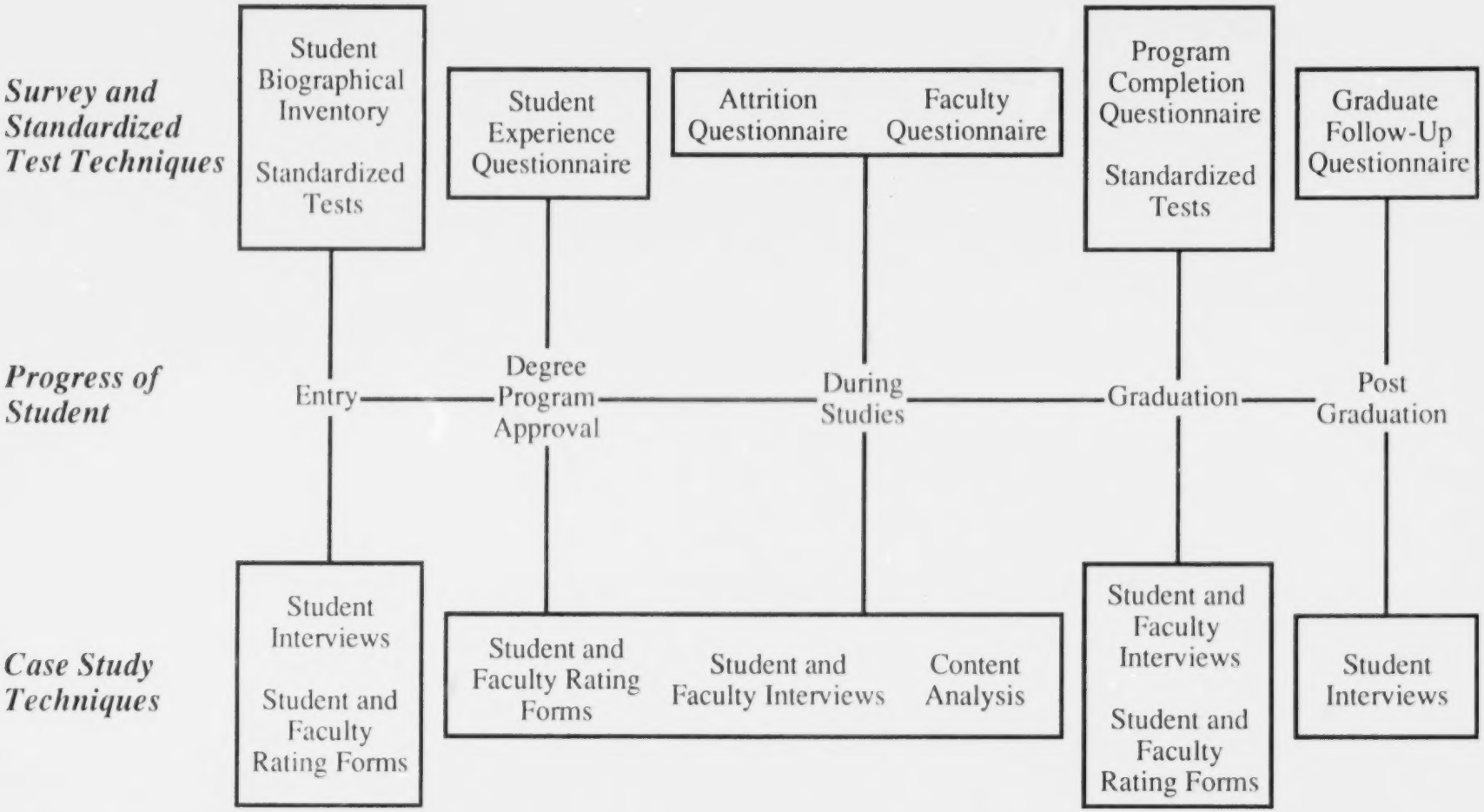
PERC's longitudinal design calls for a variety of *quantitative and qualitative approaches* to measuring student learning and change (see Figure 3, next page). A variety of surveys and case study techniques, for example, may be used in concert throughout the evaluation process. This strategy provides the necessary *multiple measures* at each stage to create *chains of evidence* showing where a college or program is having effects and where it is not (Campbell and Fiske 1959; Sieber 1974; Hartnett 1974; Alverno College - Mentkowski & Doherty 1984; Careering through College; Astin 1985, 1988; Banta 1985; Lehmann & Holtan 1988; Evell 1985).

PERC provides for both *formative and summative evaluation*, employs a variety of quantitative and qualitative evaluation methods, and focuses analyses upon adults' learning, the teaching-learning process, curricula, faculty, special programs as well as the institution as a whole. A wide variety of outcomes measures are used to assess learner goals and the cost-effectiveness of the program itself. Let us now illustrate the significance of the PERC approach by drawing upon several studies of Center for Distance Learning (CDL).

Implications of Evaluation Data: The CDL Experience

How can the seemingly complex, comprehensive, and abstract research approach just outlined help a faculty member, adult student, program administrator, or external evaluator not only understand the learning process but assess the outcomes of its academic program? Let us first discuss the Center for Distance Learning (CDL); its students, and selected findings of general interest to the distance learning reader.

Figure 3. PERC LONGITUDINAL DESIGN



Source: Palola and others, 1977.

Lehmann & Granger, Empire State College, USA / 110

The Center for Distance Learning is modeled on the Open University in Great Britain. Begun a decade after the Open University was founded, CDL serves students through academic programs and courses specifically designed for learning at a distance. Students are guided and supported by faculty and course tutors who work with them by phone, mail and personal computer.

As one of the special purpose centers of Empire State College, students in CDL can pursue entire degrees in this mode. Students served include adults whose personal and professional circumstances do not permit pursuit of a conventional program as well as those who prefer the programmatic and pedagogical flexibility of distance learning. CDL works with both the individuals and groups, adapting its content and operation appropriately. The Center uses and adapts various technology and delivery modes to distance learning, such as video, audio, and computer.

CDL offers more than 120 different courses each year and currently enrolls over 1500 students each term. In fact, CDL yearly enrollments have grown from 300 in 1979 to over 4500 in 1990. CDL's courses consist of carefully structured learning materials designed especially for adult independent study. Many of these courses are designed to enable adult students to incorporate their own learning from experience into their assignments. The course materials may include textbooks, workbooks, course guides, television programs, audio cassettes, computer conferencing, and similar resources. Students work through these course materials under the guidance of the faculty tutor who maintains regular contact with the student by phone, through the mail, or by computer.

CDL offers three areas of concentration in contrast to the eleven offered by ESC in its regional centers: Business, Human Services, and Interdisciplinary Studies. Within the Business area, a baccalaureate in Fire Service Administration is available to residents of eight New England States. CDL is designed, then, for students who are self-motivated and capable of studying independently. They prefer a program that requires no travel or classes. Students enrolling in courses at the Center for Distance Learning may be in one of three enrollment categories:

- non-matriculated student, pursuing the individual course (or courses) with no declared intention of earning a degree;
- matriculated with the CDL for an associate or baccalaureate degree, completing all or most of their studies through CDL courses;

- matriculated with an ESC regional center where they receive their primary academic advising and design *learning contracts* with a faculty member. CDL courses are often included as components of learning contracts.

Educational Planning and PROLOGUE

The purposes and needs of matriculating students are not immediately evident when they first enroll. Educational planning is undertaken by the student working with a faculty member using a specially designed course. Educational planning consists of a series of readings, exercises, and assignments to guide students through the appropriate analytical and reflective processes at a distance.

An important feature of the educational planning process now under development at CDL is an orientation and diagnostic activity called PROLOGUE. Through PROLOGUE, CDL faculty are employing an entry level diagnostic assessment of adults' learning skills and deficiencies which will enable them, both at the outset of study and over the entire degree process, to strengthen already acquired competencies and address skill weaknesses. In the words of the Educational Planning course guide, PROLOGUE activities "help set the stage and create the context" for all students' education by diagnosing, describing, and evaluating their "unique profiles of developed and undeveloped abilities." The focus here is on assessment as a natural teaching/learning process.

The ultimate intent is to design courses with a range of inbuilt learning options which can then be specified by faculty and tutors in response to the needs of different students as identified through PROLOGUE. Such a design maintains the efficiency of the structured course package and much of the effectiveness of an individually tailored study (Granger 1990; see also Benson & Dehner 1991 and Dehner 1987). It also provides for each student's academic program both an immediate and longitudinal assessment, provided by the evaluative judgements of each course tutor and the student's overall mentor.

Course Tutors

While courses are designed to be self-contained for independent study, a tutor is assigned to each course. The Center's experience has been that the course tutor is essential to most students' success in the course. A course tutor will work with up

to 35 students in a course and is responsible for maintaining an "academic dialogue" about the course material with each one through the mails and over the telephone. The tutor initiates the dialogue by contacting the student at the start of the course and then sets out a schedule of contacts for the term. Usually these contacts focus on aspects of the assignments, problems the student may have with them, or concerns the tutor may have about the student's work. The tutor's task is to support and assist the student's own learning process, and the Center devotes much attention to the quality of this support. Tutors provide students with written responses to each assignment, and the final course evaluation is expressed in a narrative rather than a letter or number grade. Every evaluative opportunity is used to guide and encourage effective learning; thus the final evaluation is formative as well as summative, also providing valuable information to the student's advisor in counselling about further study. The Center attributes its average completion rate of 70-75 % to the successfull integration of the course materials, the carefully structured term period, and the supportive guidance and pacing of the course tutor.

Faculty Involvement

American faculty who have been active in distance learning projects - either as members of a course team or as course tutors - consistently find the experience engaging and valuable. Course teams work on the principle of interdependence, bringing together content experts, frequently from several institutions, with instructional designers and media specialists to sort through the best modes and manner of content presentation. Wrangles focus on both academic and pedagogic issues, so that the course team becomes a "teaching community" intent on fully engaging a range of learners. Content experts not only must mesh with other content experts, but they are exposed to quite different perspectives on content and presentation from the other team members.

The course tutor is in many ways the most critical member of this teaching community (though often not known to the original course team), linking the prepared materials to individual real students. The tutor differs from the classroom instructor in that he or she is not a primary source of information, but rather the expert guide to the course terrain, exploring it with the student. Consequently, the tutor can make new or additional connections to the content for the student and can re-route the student's path through the course for particular purposes. Likewise, the student can illuminate the material from his or her own perspective.

As one distance tutor put it, "I know I have been successful when my students become my teachers."

Because the course content is presented primarily in the materials, the tutor can make these critical conjunctions fairly efficiently, often over the phone or through the mail, electronic or otherwise. Distance learning course teams now are developing courses with maximum flexibility built in so that the course tutor can individualize the material more readily (Granger 1990, p. 50).

Contrasting Demographics: Empire State and Distant Learning Students

Increasingly, educators agree that there are many roads to learning and that the diversity among learning styles and backgrounds of learners themselves should be reflected in the learning modes available. There is a growing interest in individualizing instruction and shaping both content and pedagogy to the needs of individual students.

Student Diversity at Empire State College

Demographically, adult learners fit the profile for diversity. Two-thirds of the 7100 enrolled students are between 30 and 50 years old, with an average age of 37 and an age range from 18 to 80. Over 64 % are women; two-thirds are married. Almost one-fifth (18 %) identify themselves as minorities. Economically, students are upwardly mobile and employed in semi-professional, supervisory, technical and clerical positions, earning an average salary of $ 17,500. One-third of the adults, however, have incomes below $ 10,000, and one-third must have financial assistance to complete a degree. Although ESC students come from all 62 counties in New York State, from 25 other states, and from at least 10 foreign countries, ninety-nine percent of the regional center students come from New York State. Eighty-five percent are employed (70 % work full-time) and study part-time (84 %). In terms of prior education, 20 % have only a high school diploma, 40 % have some college experience, 31 % already have earned an associate's degree, and another 9 % hold a college or graduate level degree. Adult learners aspire to further education, as our surveys show almost 60 % desire to pursue a masters or a professional/graduate degree. (Office of Research and Evaluation, *The Adult Learner in Nontraditional Education*, 1986, and *Focus Report on Outcomes*, 1989)

The primary goals of adult learners at Empire State include (in order of priority): increasing knowledge of an academic field, attaining specific skills useful on a job, enriching their own learning and personal fulfillment, and improving their professional status. Business, Management, and Economics is the most popular area of study, followed by Community and Human Services, Science, Math and Technology and the Arts. Students see the opportunity to work and study at the same time and to earn credit for prior college level learning as especially appealing features. ESC adult learners view themselves as strong on independence, persistence, drive to achieve, leadership ability, and academic ability - traits characteristic of adult students (Office of Research and Evaluation, 1986, pp. 8-9). On the basis of demographic diversity, Empire State College is different from many external degree or adult education programs because it attracts and seeks to serve a real cross-section of the population. With more than 17,000 graduates in 19 years, Empire State exemplifies an effective means for meeting the educational needs of adult learners in New York State.

CDL Students

The profile of CDL distance learners is similar to the general ESC profile but differs in these important ways: 60 % are men; only 11 % are minority; 60 % come from New York State and the remainder come from 22 other states and from foreign countries; almost one-third have incomes over $ 50,000 and the median income is $ 35,000; 70 % are married; and 52 % seek to obtain a master's or higher degree (CDL Survey, January 1990).

The primary goals of distance learners (in order of priority) include: attain specific skills useful on a job; improve their professional status; enrich their own learning; improve their leadership skills; and improve their chances of earning more money. Three quarters of the distance learners were studying Business (including 23 % in Fire Service Administration), followed by Community and Human Services and Interdisciplinary Studies. Distance learners see the flexible scheduling of academic work (75 %) and the opportunity to work and study off campus at the same time (67 %) as especially appealing features of CDL. CDL adults also cite independence allowed by distance study (65 %) as a very important reason for enrolling. Distance learners view themselves as strong on independence, drive to achieve, persistence, ability to handle stress, and resourcefulness. CDL students rated themselves above average on these academic skills: studying/learning "on your own;" learning from books; and learning with only occasional contact with faculty or tutor (drawn from 24 item skills list). Distance

learners thus see themselves as well prepared for distance study and represent, demographically and attitudinally, a somewhat distinct subpopulation of ESC students (CDL Student Survey, January 1990).

Distance Learners' Views of Communications Technologies

In Spring 1988, Roberts surveyed all the enrolled students in CDL to ascertain the levels of student interest in and access to various communications technologies. Roberts found that "nearly" all students have access to audio and video cassette equipment and are interested in taking courses using these technologies. Approximately 60 % have access to computers (17 % have access to computer with modem). Over 50 % of the respondents would take courses using software and over one-third courses requiring computer applications. Few respondents have access to video disc players or personal satellite downlinks. However, over a third would attend satellite telecourses at an institutional site (Roberts 1989, p. iii).

Some of the more intriguing group differences from this survey include: males are more likely to own a computer than females; two adult households are twice as likely to own a computer as one adult households, whether there are children in the household or not; males are more likely to want to use computers, software, and interactive programs in courses than females; and the age group 24-33 is most interested in viewing televised programs, whereas those over age 53 are least interested. One of the most surprising findings was that distance learners, compared with regular ESC students, were significantly less interested in renting course videotapes. Two other significant variations were: females expressed greater interest in direct off-air viewing of courses than males and respondents residing in New York State were more willing to take a broadcast telecourse than out-of-staters (Roberts 1989, pp. 14-20).

In contrast to some previous research (Heermann 1988 and Bissell 1986), over 50 % of the adult learners said they would take a course that required using software. Roberts found that "interest in using specific computer applications exceeded ownership of computers in every case and by as much as 44 % for access to library catalogs. This finding may be interpreted as indicating that a significant factor in student interest is utility. In other words, students are willing to become familiar with computers and to acquire access to them, if their application is perceived as useful to their studies (Roberts 1989, p. 26).

As a result of this survey based on 366 respondents, Roberts recommended to the CDL administration that the center increase its video and audio mediated course offerings immediately. Further, it was recommended that computer mediated courses continue to be developed, but on a more gradual and staged basis to build interest. For an example of computer conferencing in CDL courses, see Roberts (1988). Finally, Roberts recommended that CDL should market its increased use of technologies to attract a new student clientele, the technologically oriented (Roberts 1989, p. iii).

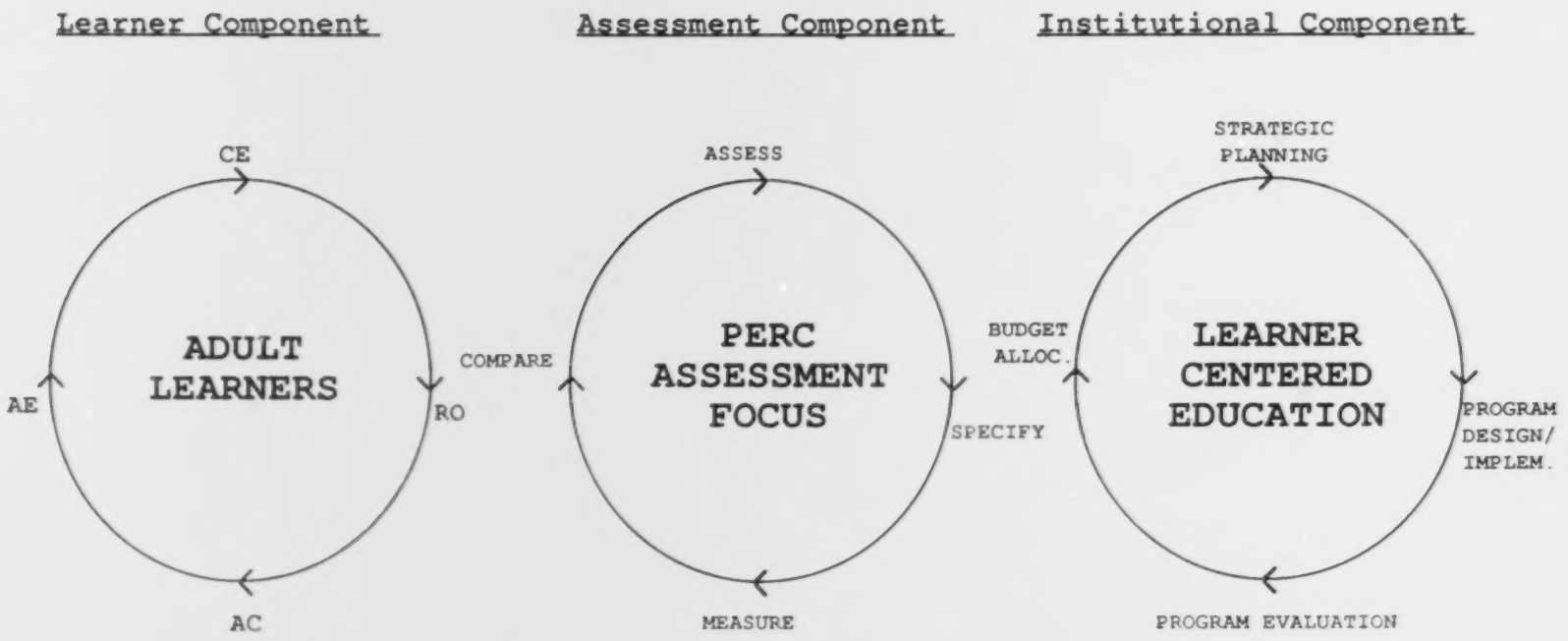
Assessment and College Decision Making

The relationship among adult learning (or distance learning in this case), assessment, and college decision-making can be shown in the diagram on the following page (Figure 4). As a pioneering college designing learner-centered programs for meeting adults' educational, career, and personal goals, Empire State College has introduced an important model into the world of higher education.

Starting with the adult learners circle, the ideas of experiential education developed, for example, by David Kolb (1981, 1984), can be used as an illustrative lens for better understanding the basic process of learner-centered education. Most adults come to ESC with concrete practical goals in mind (Concrete Experience), then pursue the design of a degree and complete a portfolio assessment process which makes them reflect on their prior learnings and conceptualize their present situations (stages of Reflective Observation and Abstract Conceptualization). Finally, they earn a degree through a series of learning contracts (Active Experimentation). The Kolb experiential learning model seems to be a particularly useful model for ESC because it is based on similar assumptions of experiential education found at the College. Further, it seems to reflect the process that adult learners go through at the College as they design their degrees and carry out their contract studies.

At the institutional level, the College has been engaged since 1980 in a strategic planning process which includes a program design and implementation phase, a program evaluation phase, and a budget allocation phase. This decision making model provides a useful way to incorporate assessment activities into the process whereby the College makes decisions about program improvement and allocates funds to support existing programs and new ventures (Lehmann 1981).

Figure 4. DIAGRAM OF ESC MODEL



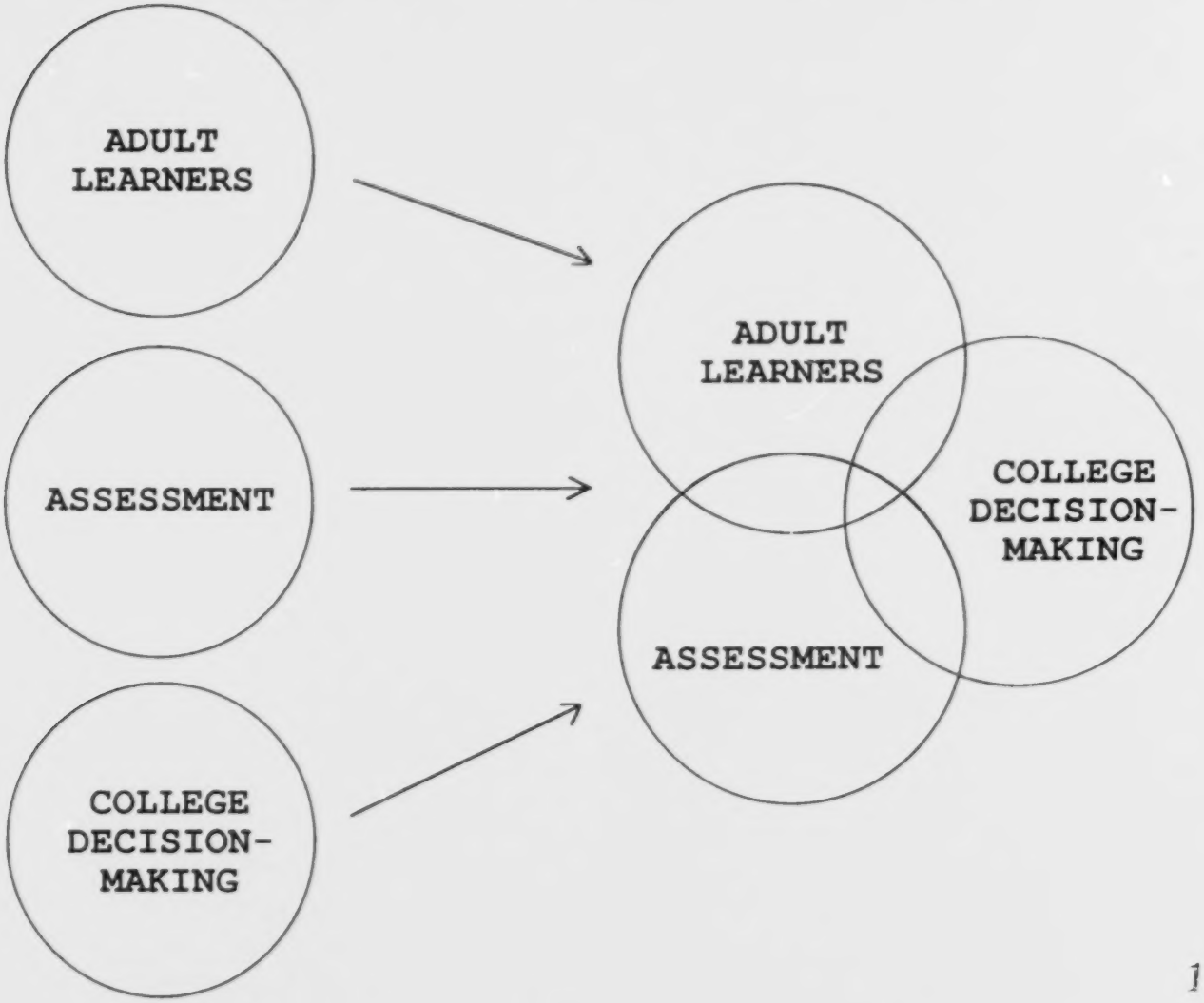
Source: Lehmann, 1989.

Connecting the adult learner circle with the decision making circle is the assessment/research circle. Although adult learners are treated individually by the College, the aggregate profile of individual cases can be analyzed with the PERC model. Program evaluation requires assessment strategies that link regional and special purpose centers, degree programs, areas of study, and concentrations with individual student learning. The assessment or research circle thus becomes a most important bridge between the adult learner and the entire college as student learning and program impacts are considered. Over time, these three separate circles begin to merge and evolve into a powerful model for assessing outcomes, measuring student learning, and facilitating more effective decision making (see Figure 5, next page). The model assumes a continuous strategic planning process, a longitudinal data base, and case studies of individual learners that demonstrate eductional outcomes in support of the quantitative assessments provided in the program evaluation cycle.

Conclusion

Empire State's assessment efforts have been developed out of the College's mission and over the years have become an integral part of the way the College serves adult learners and makes academic decisions. The evaluation model (PERC) described in the first part of this article is a comprehensive, complex, cost/effective and collegial effort to assess and compare results on adult learning in an alternative setting. There is a major focus on formative evaluation which is strongly embedded in CDL's PROLOGUE activities: Results from evaluation studies of CDL's students, graduates, leavers, curriculum, faculty, and delivery system have been used by both administrators and faculty to improve the program and enhance its effectiveness. A sound assessment strategy such as PERC can be a most important foundation for demonstrating the effects of distance education to a variety of audiences often skeptical of such alternatives.

Figure 5. ESC MODEL OVER TIME



Source: Lehmann, 1989.

Lehmann & Granger, Empire State College, USA / 120

References

Astin, A. W. (1985): *Achieving educational excellence*. San Francisco: Jossey-Bass .

Banta, T. W. (1985): Use of outcomes information at the University of Tennesse, Knoxville. In: Ewell, P.T. (ed.): *Assessing educational outcomes*. San Francisco: Jossey-Bass, Inc. New Directions for Institutional Research, No 47, 1985, pp. 19-33.

Benson, L. & Dehner, Th. (forthcoming, 1991): Close planning with distance students: The Prologue Program of the Center for Distance Learning. *Open Learning*.

Bissell, S.A., et. al. (1986): *Distance teaching techniques using electronic conferencing*. Unpublished paper. Rochester, NY.

Campbell, D. & Fiske, D. (1959): Convergent and discriminant validation by the multitrait-multimethod matrix. *Psychological Bulletin*, Vol. 5, pp. 81-105.

Center for Distance Learning (1988): *Center self-study Prepared for the Middle States Reaccreditation Review*. Saratoga Springs, NY: Empire State College.

Center for Distance Learning (1990): *College programs for adult learners*. Saratoga Springs, NY: Empire State College.

Center for Distance Learning (1990): *1990-1991 Course Catalog*. Saratoga Springs, NY: Empire State College.

Center for Distance Learning (1990): *Results from the Distance Learning Survey*. Saratoga Springs, NY: Office of Research and Evaluation, Empire State College.

Center for Distance Learning (1990): *Report of the Curriculum Survey by ESC Faculty*. Saratoga Springs, NY: Office of Research and Evaluation, Empire State College.

Dehner, Th. (1987): Prologue at the Center for Distance Learning: A skills assessment strategy for distance learning. Paper presented at *Third Annual Conference on Teaching at a Distance*, University of Wisconsin at Madison, August 1987.

Empire State College (1979): *Access, Innovation and the question for excellence: Empire State College in 1979*. Prepared for the Commission on Higher Education, Middle States Association of Colleges and Schools. Saratoga Springs, NY: Office of the President.

Empire State College (1989): *Focus report on outcomes*. Saratoga Springs, NY: Office of Research and Evaluation, January 1989.

Empire State College (1989): *Retrospect and prospect: Strengthening learning at Empire State College*. Prepared for the Commission on Higher Eduction, Middle States Association of Colleges and Schools. Saratoga Springs, NY: Office of the President.

Empire State College (1990): *Plan for outcomes assessment*. Saratoga Springs, NY: Office of Academic Affairs, June 1990.

Empire State College (1989): *Reports and Publications 1979-1989*. An annotated list of reports prepared by the Office of Research and Evaluation, October 1989.

Ewell, P. T., (ed.) (1985): *Assessing educational outcomes*. San Francisco: Jossey-Bass, Inc. New Directions for Institutional Research, No. 47, 1985.

Fong, B. (1987): *The external examiner approach to assessment*. Paper prepared for the Association of American College, Washington, D.C..

Granger, D. (1986): Empire State College's tradition of innovation: The Center for Distance Learning. *ICDE Bulletin*, Vol. 10, (January 1986), pp. 24-29.

Granger, D. (1990): Open learning and individualized distance learning at Empire State College. *Open Learning*, Vol. 5, No. 1.

Granger, D. (1990): Open universities: closing the distances to learning. *Change*, Vol. 22, No. 4 (July/August 1990), pp. 44-50.

Granger, D. & Roberts, L. (1989): Individualization in distance learning: How far do we go? *Research in Distance Education*, Vol. 1, No. 1 (April 1989), pp. 8-9.

Hartnett, R. (1974): *Problems with the comparative assessment of student outcomes in higher education*. Princeton, NJ: Educational Testing Service.

Heermann, B. (1988): *Teaching and learning with computers*. San Francisco: Jossey-Bass.

Hodgkinson, H. (1975): Evaluating individualized learning. In: Berte, N. (ed.): *Individualizing education by learning contracts*. Pp. 75-84. San Francisco: Jossey-Bass.

Lehmann, T. (1989): *Assessment at Empire State College*. Prepared for Submission to SUNY Central Administration. Saratoga Springs, NY: Office of Research and Evaluation, May 1989.

Lehmann, T. (1981): Evaluating adult learning and program costs. In: Chickering, A. and Associates: *The Modern American College*. Pp. 748-772. San Francisco: Jossey-Bass.

Lehmann, T. (1988): Fulfilling democracy's promise through education: The Empire State College experiment. *Golden Hill Education Issue*, Vol. 4, pp. 5-45. Saratoga Springs, NY: Empire State College.

Lehmann, T. & Holtan, J. M. (1988): *Accounting for quality through assessment*. Albany, NY: Rockefeller Institute of Government and Empire State College, SUNY.

Mentkowski, M. & Doherty, A. (1984): Abilities that last a lifetime: Outcomes of the Alverno experience. *AAHE Bulletin*, February 1984, pp. 5-14.

Paloia, E. G., Lehmann, T., Bradley, [illegible] & Debus, R. (1977): *The PERC Handbook*. Saratoga Springs, NY: Office of Research and Evaluation, Empire State College.

Roberts, L. (1985): *Computer conferencing: A classroom for distance learning*. Bulletin, International Council for Distance Education, Vol. 18, September 1985, pp. 35-40.

Roberts, L. (1988): *A comparison of the reasons adult and traditional age students attend a nontraditional college*. Practicum paper presented at Nova University, Ft. Lauderdale, Florida, November 1988.

Roberts, L. (1989): *Analysis of student interest in and access to distance applications of advanced eductional technology*. Practicum paper presented to Nova University, Ft. Lauderdale, Florida, January 1989.

Roberts, L. (1989): *Analysis of the educational objectives of adult learners*. Practicum paper presented to Nova University, Ft. Lauderdale, Florida, February 1989.

Sieber, S. D. (1973): The integration of fieldwork and survey methods. *American Journal of Sociology*, Vol. 78, pp. 1335-1359.

Evaluation of Danish Distance Education at University Level

Annette Lorentsen,
Aalborg University, Denmark

Abstract

Evaluation of distance education in Denmark centres around specific cases as there is no permanent central evaluating authority.
As what and how is evaluated depends to a high degree on the context evaluated, I shall by way of introduction explain the Danish distance education concept which differs from the internationally dominating British Open University concept. Subsequently a more detailed account of the three largest evaluation cases in Danish distance education is given. Summing up, the general characteristics of these evaluation activities are gathered in a total picture of a general philosophy of evaluation within Danish distance education.

Evaluation is the evaluation of something carried out in a specific way for a specific purpose. Contents, form and purpose are therefore very important parameters in any discussion of evaluation activities (Thorpe 1988).

Before we are able to analyse and discuss these levels of given evaluation activities there is, however, an even higher level to be clarified. A fruitful engagement in evaluation activities at any level - whether you carry out the activities yourself, analyse other people's evaluation activities from the outside, or read other people's analyses of still other people's evaluations - presupposes that everybody, be they evaluators, analysts or readers, possess profound knowledge of the conceptual ideas and the characteristics of the context within which evaluation is carried out.

If for instance we talk about the evaluation of specific elements of teaching, it is necessary to define what we mean by teaching. Is teaching synonymous with the transfer of knowledge from teacher to students, or do we understand teaching as a common production of knowledge (Fox 1983)? Such defining clarification sets the frames of what is relevant to evaluate (i.e. the *design* of the evaluation), and how the results of the evaluation should be *interpreted*.

Teaching according to the transfer model does not make it nearly as relevant to evaluate the processes among the students themselves as does teaching based on the theory that it is common production of knowledge between teacher and student and among the students themselves.

Furthermore, in the former case many contacts among the students during the period of teaching will be interpreted as a high degree of noise - in the latter, however, as a form of a measure of success as to whether the common knowledge-creating processes have started.

To sum up one could say that the evaluator or the reader of evaluations must know the whole set up being evaluated very well in order to formulate appropriate questions or focus on appropriate issues within an evaluation.

When Thorpe (1988) started dealing with the evaluation of open learning and distance education without explaining in detail the concepts behind open learning and distance education, she could do so due to the existence of a common knowledge of the Open University's open learning and distance education model which - as indicated on the back of her book - is based on course 'production' and course 'delivery' with course material and learners as the motors of the course, between which tutors and counsellors intermediate.

If your evaluation frame is unknown or differs from the prevalent understanding of specific concepts, you must, however, be explicit to ensure the establishment of a common frame of understanding of the subsequent evaluation activities. This is the case with Danish distance education. In Denmark there are not the same principles and ideas behind open learning/distance education as there are traditionally in other countries.

Consequently, it is necessary to clarify these fundamental principles and ideas on which Danish distance education is based. For instance it is vital, for an understanding of Danish distance education, to know that Jutland Open University (JOU) is not an independent institution in a traditional sense with its own policy, its own teachers and students etc., but a cooperation enterprise concerning decentralized distance education activities at several universities, which means that the concrete distance courses have to stick to the rules and the concepts of 'a mother university'.

It is also vital to know that Danish distance education is not divided into a dichotomy of course production and course delivery, which again means that the common concepts of course production teams and tutors do not reflect a Danish context.

Therefore, the following contains a discussion of a few of the most central characteristics of Danish distance education at university level before concrete evaluation activities related to this context are analysed.

Distance Education in Denmark within the Open University

In order to get a fair impression of what the open university is like in Denmark I suggest we look at three issues: educational policy, pedagogy, and organization of the Danish open university activities.

Educational policy related to the open university in Denmark

Like the Open University in Britain, the Danish open university activities were started in order to broaden the access to higher education. Therefore, no formal qualifications were obligatory to enter the open university. And - a very important factor in Denmark - education was free, as Danish higher education traditionally is.

However, right from the start of the Danish open university it was stressed that it was not sufficient to aim at this formal access to higher education, although this was, of course, an important element in the democratization of education. The preparation of the conditions of study for the future students was equally important if open education was to be a real offer to adults.

This problem has been realized by other open universities too, but it is characteristic that the answer to the problem is different in the Anglo-Saxon tradition and in Denmark. While the British Open University tried to create appropriate study conditions for adults through an orientation towards the *individual* and by making the studies as *flexible* as possible for the individual student in relation to his domicile, job, family and other obligations - i.e. created *individually flexible frames*, the founders of the Danish open university rather regarded it a *pedagogic* problem to make the formal access to university real to adults:

> "If we want to open up the spectrum of education towards other groups of receivers than the traditional groups, it is necessary to define the education in relation to the conditions of learning for students who are less mobile geographically and as regards time - but who are more open in their motivation for learning and cognition than traditional students.
> Consequently, the open university must be an organization which through distance and part-time studies, pedagogical new thinking, use of new information technologies, and a strong support from the individual institutions, opens up for an active acquisition of knowledge by new types of students. ( )
> The open university education will make untraditional demands to the institutions as regards the planning of the education. The development of new, open forms of study should take place simultaneously with the formal opening of the institutions. The necessary flexibility in the adult education offered by the institutions can only be established through such development work."
> (Åben Universitetsundervisning i Danmark 1985, p. 15 and 27)

Since the establishment of the open university activities in Denmark there have, therefore, been experiments aiming at creating a pedagogy around distance education for adults. The creation of such special pedagogy has been based on the overall Danish tradition for research in adult pedagogy.

Open university pedagogy in Denmark

A characteristic feature of the open university pedagogy in Denmark is the fact that it focusses on *communication* as the central element in education, both communication between teacher and students and among the students themselves. This view of the central issue in distance education was new at the time when the Danish open university was established. Today there are tendencies showing that it is finding favour. This at least is the opinion of Garrison (1989), who says:

> "The focus of distance education is shifting from structural issues to process issues. How we facilitate mediated communication will guide the development of distance education in the future."
> (Garrison 1989, preface)

The Danish open university pedagogy represents, therefore, a paradigm quite different from the traditional paradigm within an open university.

> "Historically, distance education has placed great emphasis on individual and independent study. The educational process consisted largely of students 'interacting' with print materials. Mediation and recognition of learning was minimal. ( )
> In distance education greater efforts generally are put into preparing or packaging contents (i.e. structure) and much less effort is given to the crucial element of the educational transaction (i.e. the interactive dialogue with the purpose of negotiating objectives and pursuing meanings)."
> (Garrison 1989, pp. 9 and 19)

The Danish concept might be expressed as a *communicative distance education concept* with focus on learning (creation of knowledge) through *interaction* and *communication*.

In contrast to this stands the traditional *knowledge distributing distance teaching concept* - represented by, for instance, the large European open universities - "focussing on teaching viewed as distribution of knowledge to individual students seen as empty vessels" (Lorentsen/Rasmussen 1989, p. 9).

The Danish open university education is, to use a phrase from Garrison (1989), "a socially recognised activity that is realised through interaction - not independence" where "communication is the interface between teaching and learning" (p. 122).

Organization of the open university in Denmark

As regards organization there are, too, several points where the Danish system differs from the models of other countries.

First of all, in Denmark the open university is *de-centrally* organized, which means that it is organized in the way that all existing ordinary universities are, offering both traditional education and, to a much more limited extent, distance education (dual mode institutions). Jutland Open University, which is the oldest and largest open university in Denmark, is therefore no independent institution, but exclusively a joint venture between the faculties of the humanities of primarily two existing universities.

All courses under JOU are offered by one of the universities involved in accordance with the rules of study structure and contents which apply at the institution in question.

As regards the development of the individual courses the idea of decentralization is carried through as well. So far there have been no central course development teams at the open university in Denmark. The individual teachers offering a course have defined, prepared, and administered the instruction material regarded as being most relevant for the course in question. The teachers are autonomous as they are at the traditional university in Denmark - both as regards choice of material and choice of teaching method.

> "One might say that the Danish system of distance education at university level tries to maintain the educational structures of the 'mother-university'. Thus, not only are the contents of the courses the same as for the on-campus courses, but so, too, are the teachers' functions (although there are some additional functions when dealing with distance education). It is the teachers who produce or collect the course material, it is the teachers who do the tutoring and counselling, and it is the teachers who correct papers and arrange examinations. Unlike the British Open University there is therefore not a division of labour between a course team and the tutors. It is then a form of distance education that has more in common with on-campus education than with the industrialised form of distance education described by Peters (1973)."
> (Christensen 1990)

This, however, is changing a little in Denmark now - not towards large course development teams isolated from the concrete education, but still with a tendency for new projects within open education to let a certain amount of educational material pre-produced jointly by the whole team of future teachers.

Within the open university in Denmark teaching is always organized around classes of students, in practice groups of about 30-40 students who will be together all the way through their studies.

This means that there is a frame around the activities of the students offering the possibility of subject-related as well as social contacts with like-minded students. Hereby Danish open university activities differ clearly from the dominating international orientation towards the individual in distance education.

Having studied the characteristics of the Danish open university the frames for a relevant evaluation of the activities in question have been set, for it is obvious that the distance education concept chosen will decide on which issues the evaluation should focus in order to catch what is crucial within a given concept.

Below we will look at the three largest *evaluation cases* within distance education in Denmark. After dealing with the individual cases, some general characteristics of evaluation activities within Danish distance education at university level will be discussed and the status and future perspectives of such an evaluation will be touched upon.

Case 1 An evaluation of the first 5 years of Jutland Open University's existence 1982-1987 (Hald 1987).

After five years of existence Jutland Open University made a general evaluation of its activities so far. The evaluation was carried out by a team of two persons who were working temporarily at JOU.
The description here is founded on the final evaluation report (Hald 1987).

Purpose and areas of the evaluation

In Hald (1987, p. 9) it says that the purpose of the evaluation is to give a general picture of Jutland Open University's activities from 1982-1987. All the way through the evaluation this general picture is constantly confronted with the original educational political purpose for founding Jutland Open University. Therefore, the purpose of the evaluation may be summed up as follows:

> The aims ( ) are to give a description of ( ) JOU and to investigate to what degree JOU has been able to fulfil its purpose, e.g. to contribute to a more equal access to ( ) and to remove the traditional barriers to higher education.
> (Hald 1987, p. 9 and p. 31)

It is not the idea of the evaluation to evaluate individual courses, but rather to evaluate whether the pedagogical concept of the institution JOU and the pedagogical practices at JOU encourage a broader recruitment to higher education in Denmark, which JOU made into its goal from the very beginning.

Data

The evaluation is founded on four types of data:

- statistical administrative data available at JOU (How many courses and what courses have been run? Which students have taken part in which courses? etc.)
- data taken from the students' applications for starting their courses
- survey among all students and teachers between 1982 and 1986 - including drop outs
- interviews with students.

As can be seen these data represent a *mixture of quantitative and qualitative data*.

Outcome variables

The output of this evaluation is divided into the following variables:

- subjects offered (number and type) and number of applicants

- profile of the students attending the courses during the first five years at JOU through description of:
  - domicile
  - age
  - former education
  - occupation

- profile of the study processes through description of:
  - general aspects linked to the planning and the realization of the courses, such as for instance: the teachers' opinion of the students' qualifications compared with the qualifications of the full-time students; the distance students' aim and motivation for studying; description of to what extent the distance students are involved in the planning of the teaching; the students' evaluation of the teaching; the importance of social aspects to learning;
  - aspects linked to the residential weekend seminars which form a very important part of Danish distance education courses, such as for instance: practical problems connected with these seminars; how teachers collaborate

in connection with the seminars; how the students are organized during and after their first seminar (whether local study groups are established by the teacher, for instance); teaching methods at the seminars;
- aspects linked to the home study periods, such as for instance: what material is used by the teacher; evaluation of the teacher/student relationship; ways of studying during these periods (individual work, group work); teachers contact to students during these periods.

- description of study results:
  - completion rate.
  - drop out rate.

Findings of the evaluation

The final conclusion of the evaluation is that JOU has contributed to spreading higher education in Denmark. JOU has been able to reach people in jobs with a higher degree of spreading as regards geography, age, and sex than the traditional types of education. However, many of the distance students are already educated people.

Even more important than these results related to the original educational political issues of the evaluation is the fact that the evaluation very clearly reveals the pedagogical strengths and weaknesses of the JOU model.

A strength is no doubt the residential weekend seminars which are praised to the skies by both teachers and students, because here an incredible intensity and dynamics are achieved which further both the study processes of the distance students, the social contact between students and teacher and among the students themselves, and the teachers' understanding of the needs, etc. of the students.

The weaknesses of JOU's pedagogical practice, however, appear clearly, too. They are to a high degree concentrated on the home study periods between the very popular and fruitful residential seminars. Two main problem areas are located as to the home study periods, i.e. students lack contact to other students and to their teachers during these periods, and students want more *teacher activity* (support, teaching, materials etc.) during these periods.

Summing up it may be argued that this evaluation as its starting point asked some very general educational-political questions which were answered broadly, but its most important result, however, turned out to be different, as it rendered visible some problems of the institution's pedagogical practice concentrated

around the lack of contact and the lack of teacher visibility during the home study periods.

The two following cases may be looked upon as different answers to Hald's demonstration of a need for better home study periods within Jutland Open University.

Case 2 is an attempt by the University of Copenhagen to qualify the home study periods through use of television and broadcast - both as traditional one way television and broadcast programs and as telephone programs containing the possibility of posing questions, raising issues, getting discussions started etc.

Case 3 is even more directly related to Hald's findings within Jutland Open University, since case 3 contains Jutland Open University's answer to its own evaluation. To Jutland Open University the answer to Hald's description of problems in the home study periods is the establishment of a communication medium in order to meet the stated need for contact and communication.

Case 2 Evaluation of a media-supported course in music history

At the end of 1987 and during the first half of 1988 the University of Copenhagen in cooperation with Danmarks Radio (the Danish broadcasting corporation) carried out an open university course in music history. The experiment was evaluated by an employee from each of the two institutions (Sederberg & Anker Olsen 1989).

In its pedagogic planning the music history course reminds us in some points of the courses of Jutland Open University (seminars; telephone contact with teachers), but in other points it differs, as it pulls threads back to more traditional correspondence teaching (by using to a large extent pre-produced written material), and tests the use of new media in the form of TV and radio in support of distance teaching.

As regards educational policy, the distance courses at the University of Copenhagen as well as the JOU courses intend to reach a broader receiver group than traditional university courses.

Concepts and philosophy of the evaluation

The two researchers who have evaluated the music experiment give in their evaluation report an explicit account of their concept of what evaluation is, how it should be carried through, what should be evaluated by whom, for whom, for what reason, etc. The evaluation report, therefore, forms a good basis for encircling the two researchers' concepts and philosophy of evaluation.

Søderberg & Anker Olsen regard themselves, as they express it, as part of a change of paradigm within evaluation research (Søderberg & Anker Olsen 1989, Appendix 1 p. 1). Where previously evaluation research primarily focussed on whether the *goal* of the education was fulfilled, which was analysed via *product* analyses based on *quantitative* data, the evaluation research (including that of Søderberg & Anker Olsen) has today changed towards being more *process* oriented:

> "At the same time the view as regards how to make an evaluation has changed towards a more complex understanding of the circumstances affecting the education as a process.
> Today we endeavour to regard the educational process as a whole, and to examine which circumstances have a furthering and limiting effect on the educational process and on the results of the education."
> (Søderberg & Anker Olsen 1989, Appendix 1 p. 1)

Søderberg & Anker Olsen (1989) sum up their ideas by maintaining a change of paradigm from an *efficiency* paradigm to a *process* paradigm.

In consequence of the fact that the evaluators regard themselves as part of this process paradigm they define evaluation explicitly as follows:

> "In the following we will use the concept evaluation about a systematic analysis of the planning, the process, and the results of the open university study. This evaluation is made taking into consideration a large number of factors all contributing to an understanding of the study as a whole."
> (Søderberg & Anker Olsen 1989, Appendix 1 p. 4)

And the parameters selected for the evaluation correspond as the following is evaluated:

- the frames (including student and teacher profiles)
- selection of contents and organising of the course (i.e. study material, working forms, etc.) (the dimension of planning, see quotation above)
- the actual development of the teaching processes (the dimension of development)
- the results on the individual level (the dimension of the result).

Methodically the two evaluators explain their conscious aiming at a combination of *formative and summative evaluation* based on the different outputs of these two types of evaluation (ibid., p. 4f.):

> "Typically, the results of a summative evaluation will appear a long time after the production is finished, for which reason they are of greatest interest to the external decision-makers. Formative evaluation, however, may assist in guiding the process and avoiding striking mistakes during the preparation of the course."
> (ibid., p. 5)

Concretely, the two evaluators have in their own words played the following roles in the evaluated experiment (ibid., p. 3 and Appendix 1, p. 6):

- they have participated in large parts of the planning including making active contribution to the pedagogic discussion of the organizing frames of the course,
- they have expressed their views in connection with the pedagogical planning and adjustment of the teaching material, including their own preparation of a student's guide for the students of the course,
- they have assisted in the counselling of potential students,
- they have currently discussed teaching activities and forms of examination at staff meetings and seminars.

In this way the evaluators have obtained a very inside relationship to the course evaluated. In their own opinion this has meant that more attention has been paid to their current critisism:

> "( ) we have experienced that the "production team" to a larger extent accept critical comments and evaluation results and consequently take the evaluation seriously when they realize that the evaluation has been carried out by persons with an insight into the conditions and the process of the production, rather than by a critical authority from outside."
> (ibid., Appendix 1 p. 7)

The two evaluators themselves deal with the fact that such close involvement may result in problems of validity and reliability of the collected data and their analysis. However, according to Søderberg & Anker Olsen, this is compensated for by the fact that the critisism of the evaluated processes is more clearly and more consequently carried through when the insight of the researchers in the total process is greater. (ibid., Appendix 1 p. 7).

Data

The evaluation is based on:

- two inquiry forms for the students of the course - one before the start of the course and one half-way through the course,
- semi-structured interviews with selected students,
- observations during the Saturday-seminars which were part of the course,
- listening to the TV and radio tapes with the broadcasts related to the course (primarily pre-produced broadcasts, but also the radio question programmes used),
- participation in staff meetings.

In the summative evaluation of the study process and of the use of media the greatest importance is attached to the quantitative data. The qualitative data are more relevant as regards the formative evaluation, which is not dealt with in detail in the final evaluation report, and as regards the aspect of cooperation and cooperation problems between the two institutions involved.

Case 3 Evaluation of the use of computer conferencing in distance education courses

The third case deals with evaluation of the use of computer conferencing in distance education by Jutland Open University. The evaluation has been carried out by the PICNIC project with domicile at Aalborg University (Lorentsen et al. 1988).

The experiment with the use of computer conferencing at Jutland Open University must be seen in relation to the results of previous evaluations of JOU courses (see a.o. case 1 above). Based on JOU's concepts of distance education as well as the institution's pedagogic practice it is possible to pre-analyse problems and needs which match expected improvements via the use of computer conferencing. This has been done by the PICNIC project during the first phase of its evaluation of the use of computer conferencing at JOU. Below is a short summary which will facilitate the subsequent detailed account of the evaluation.

When Jutland Open University was established, the educational concept was stated to be: distance education as a *social learning process* characterized by an active correlation between acquisition of knowledge and production of knowledge,

therefore giving the teaching *process* priority over the development of course material (Lorentsen 1989 p. 18). Accordingly, teaching and learning should to a large extent be based on a high degree of *involvement* by the individual participant (for instance through student involvement in decision-making as to course contents, thereby rooting learning in students' experience) and on a high degree of *interaction* between students and teachers and among the students themselves both in the form of subject-related two-way communication with the dialogue as an important method of learning and in the form of social interaction as support a.o. to keep up motivation and tackle differences in student backgrounds and study-related insecurity (Lorentsen 1989, pp. 21ff).

The *key words* in order to meet these requirements, which according to prior evaluations (see case 1 above) had not taken place to the necessary extent, would both in accordance with JOU's basic ideas and the results from the previous evaluations be: *contact, mutual support, dialogue and cooperation*, which all presuppose the possibility for the students and teachers to communicate regularly with each other during the periods of self study. And that is where computer conferencing comes in, which can be seen from sources where JOU formulates its expectations with regard to computer conferencing.

In other words, JOU expects that computer conferencing will be able to strenghten *communication*, *cooperation* and *social interaction* (Lorentsen 1989, pp. 28ff). The teachers of the courses evaluated by PICNIC concretize these expectations by formulating the following demands with regard to the computer conferencing system: Computer conferencing must strengthen group work as to the obligatory project work in the course; it must improve the level of discussion in relation to both the material introduced by the teacher and the students' individual assignments; it must make teachers more visible and easy to reach; it must support student involvement and influence on their own education; and it must encourage a better relationship (create a synergy) between professional and social dimensions of the learning processes (Lorentsen 1989, p. 81).

The evaluation process

As mentioned above the PICNIC project has evaluated this use of computer conferencing at JOU. The PICNIC project is a large research project dealing with the use of computer conference systems in distance education of which the educational processes evaluated only form a part. Hereby the empirical research gains a double function, partly the traditional - more practical - function of the institution which via the evaluation can improve its courses, and partly a more scientific function:

> "The aim is not to use the empirical research findings illustratively, i.e. as exemplification of the theories used. On the contrary, it should function **exploratively**, as a corrective of the theoretical basis as well as the problem areas and analysis models presented, with a view to qualifying the latter."
> (Dirckinck-Holmfeld et al. 1988, p. 27)

Accordingly, PICNIC intends using its empirical findings in a cyclic interplay with formation and qualification of a theory within the area of distance education - a theoretical basis which not only PICNIC is looking for and wants to have established (Bååth 1979, Holmberg 1986):

> "First it is crucial to stress the importance of combining evaluation of CMC with thorough theoretical and methodological studies in order to gain a deeper understanding of what is going on. In Picnic we see these three dimensions as interrelated, whereby evaluation of CMC activities to Picnic becomes a cyclic process of collecting and interpreting data on one hand and refining our theory and methodology, on the other hand."
> (Lorentsen 1989b, p. 196)

Within the frames of Picnic, defining itself as an interdisciplinary research project situated in and interacting between the humanities and the broad range of information technologies, combining approaches and ideas from technology evaluation, communication analysis, and pedagogical research (Lorentsen et al. 1988, p. 6), an explicit account is made for the methodology, concepts etc. of the evaluation. (Dirckinck-Holmfeld 1988 el al., pp. 26 ff.).

Picnic clearly formulates itself within the modern evaluation paradigm as expressed above:

> "The overall method of which we will make use in this project must ( ) regard the research process as an interwoven subject-object relationship, be able to catch the conscious aspects of the social life (and the total character of the concepts of the test persons), and finally be a general definition of the pedagogical-communicative situation of which the test persons form part."
> (Dirckinck-Holmfeld et al. 1988, p. 27)

Due to the fact that it is the use of new technology - and a special form of new technology - that is to be evaluated, the evaluation will have a final design which will be explained in detail below.

Types of Data

Picnic characterizes its method as *qualitative* based on the extent of the individual data, the techniques, the whole preparation of the evaluation as well as its cognitive interest and aim (ibid., p. 31).

However, the project is based on *qualitative* as well as *quantitative* data, because, as stressed by Picnic, both types of data in any evaluation must be regarded as an offer which - in spite of the scientific-theoretical co-significances - should be involved to the extent which is desirable with the concrete area to be investigated.

The concrete *types of data* collected:

- logging of all (non-private) texts in the computer conference system used
- logging of information about the participants' use of the conference system (final status)
- questionnaires for students, teachers, and drop-outs
- interviews with selected students
- observation during weekend seminars
- arrangement of future workshop with a regionally defined group of students.

Among such data Picnic attaches the greatest importance to the qualitative data. The quantitative data (the questionnaire data) are primarily used to obtain factual data (sex, age, background, etc.) and in a wider sense as starter and inspiration for further qualitative data collection, a.o. to assist in selecting persons for inverviews (ibid., p. 31).

Among the data collected special attention should be paid to the use of the texts of the computer conferencing system (both those produced by the participants and the statements produced by the system) as well as the arrangement of a future-workshop as new types of qualitative data.

The texts of the computer conferencing system actually support the educational activities which take place in the virtual distance education "classroom". The texts of the computer conference system make it possible for the evaluators to attend and participate in the concrete learning activities. Of course, different basic parameters for such observation apply as compared to a traditional classroom observation, inasmuch as classroom activities in the computer conferencing system are not limited by time and place, are linked to written language, and, finally, depending on the possibilities of the system, may be hidden to the observers via production of texts etc. invisible to the evaluators.

The texts of the computer conferencing system thereby become a very important qualitative source of data, the use of which is amplified in Lorentsen (1990). It cannot, however, stand alone as there are certain shortcomings (Lorentsen & Dirckinck-Holmfeld 1990):

- not all educational activities take place via the computer conference system, and some of the activities are even invisible to the evaluators (see above),
- mental and social activities which cannot be expressed in writing must, of course, be collected using other qualitative sources of data, like for instance arrangement of future workshop treated in detail in Dirckinck-Holmfeld (1990).

Roles

According to Picnic both evaluators and participants must play certain defined roles in order to ensure the optimal collection of data as regards quality.

As far as the evaluators are concerned they must possess a very deep knowledge of the context to be evaluated, and at the same time the evaluator must be free of any ties (ties of responsibility, of emotional, economical, organizatory or any other nature) to the subject to be evaluated:

> "We find it a very important prerequisite for qualitative research and for being able to judge the experiments as a whole that the researcher has an inside knowledge of the students, the planning, the teaching, etc. At the same time, however, it is the obligation of the research to abstract from the concrete case, to be critical and analytic."
> (Lorentsen & Dirkinck-Holmfeld 1989, p. 127)

Such qualifications are rarely met in one person, for which reason evaluation teams consisting of representatives from both the outside and the inside are very often found. Picnic has been organized in the same way. Picnic finds such organization of evaluation teams explicitly ideal:

> "One could rightly maintain that the internal research relations hide a number of problems, as it can be difficult to be objective about something you have prepared yourself and for which you yourself are responsible. On the other hand external evaluations are not unproblematic either. ( )
> For instance external researchers will usually have greater difficulty in acquiring an adequate knowledge of the procedures, norms and traditions which are not explicit but nevertheless often, to a larger extent than the formal basis, constitute the reality on the basis of which educators and administrators act in their daily work.
> On this basis we find an internal as well as external research organization ideal, as it has the prerequisites of close research and evaluation work."
> (Dirckinck-Holmfeld et al. 1988 p. 30)

In Picnic, in consideration of the qualitative nature of the evaluation, the traditional subject-object relationship is not applicable either as regards the relations between evaluators and participants. According to Picnic the processes investigated continuously influence on both evaluators and persons examined, which produces an interwoven subject-object relationship with no waterproof

shutters between investigators and the object of investigation (Dirckinck-Holmfeld et al. 1989, p. 26).

Use of Data

Within the traditional dichotomy of evaluation research between formative versus summative use of data Picnic is primarily *formative*, as the evaluation is made concurrently with the development of the experiment and with a view to improving the pedagogical and communicative concept (Dirckinck-Holmfeld et al. 1989 p. 29).

As the use of new technology is the object of evaluation it is important too, however, to bring up the parallel concepts used in the research of technology estimation.

Corresponding to the concepts formative and summative in the pedagogical evaluation research are within technology evaluation the concepts *pro-active* and *re-active*, which largely contain the same questions. Within more traditional media and technology evaluation research the reactive research approach is dominant.
The media and the technology are given. What should be analysed are the consequences of using them.

In contrast to such pure consequence analyses there is the pro-active approach, which attempts to enter at a far earlier stage than in the phase of the use itself of given technologies. The proactive approach attaches great importance to the initial phase of analysis in order to catch wishes and visions which may even not have to be answered by technology. If the answer is development of technology, the pro-active approach will attempt to enter into the subsequent design phase as well as the implementation phase, partly, of course, in order to follow up on the wishes and visions collected during the analysis phase, partly in order to incorporate consequence evaluations and criticism of the technology already during the development of the technology.

In relation to the above-mentioned contrasts Picnic defines itself as a primarily pro-active evaluation project (ibid., p. 29). The technology is to a certain degree given due to the computer conferencing system used, and in this respect Picnic is reactive.
However, as Picnic says (ibid., p. 29):

"The perspective of our current evaluations during the research process (the pedagogical and communicative consequences of the technology used) is, however, pro-active. The intention is to incorporate the criticism ( ) partly successively in the current framing of the JOU experiment, partly in a future use and framing of computer conferencing systems in distance education - including indication of possible pedagogical and communicative changes of the distance education concept."

General Characteristics and Status of Evaluation Activities in Denmark - Towards a General Philosophy of Evaluation within Danish Distance Education

Having studied the three main evaluation cases within Danish distance education, some common trends appear which could point towards a general philosophy of evaluation in this area. This will be summed up below under two headlines: 'Which form of evaluation should be made?' and 'What should the evaluation be used for?'

Which Form of Evaluation should be made?

The Danish evaluations dealt with all fall within what in literature is described as "the modern approach" (Holt 1985). Consequently it is a clash with the dominating traditional scientifically inspired approach:

"Evaluation research is dominated by the largely unquestioned, natural science paradigm of hypothetico-deductive methodology. This dominant paradigm assumes quantitative measurement, experimental design, and multivariate, parametric statistical analysis to be the epitome of 'good' science. ( ) By way of contrast, the alternative to the dominant hypothetico-deductive paradigm is derived from the tradition of anthropological field studies. Using the techniques of in-depth, open-ended interviewing and personal observation, the alternative paradigm relies on qualitative data, holistic analysis, and detailed description derived from close contact with the targets of study. The hypothetico-deductive, natural science paradigm aims at prediction of social phenomena; the holistic-inductive, anthropological paradigm aims at understanding of social phenomena."
(Patton 1980, p. 19)

As appears from both Patton (1980) and Holt (1985) an evaluation concept like the Danish one is characterized by larger broadness than the traditional scientifically inspired approach:

> "More generally ,the modern approaches focus on all three aspects of learning - contexts, processes and products - and thus provide a broader conceptualisation of the evaluation task compared to the traditional approach."
> (Holt 1985, p. 10)

The characteristics of the Danish evaluation model can be summed up as follows:

- Evaluation is to a large extent *process oriented* rather than centered around analyses of outcome.
- Evaluation is designed and carried through in a *flexible* way which makes it changeable in case unforeseen circumstances should so require.
- The participants evaluated are involved in the evaluation process instead of being regarded as objects from the outside.
- *Qualitative data* play an important role (e.g. observation data and other types of qualitative data primarily known from the social sciences, where they are used by , for instance, sociologists, ethnographers, anthropologists and psychologists).
- The totally neutral, objective evaluator is regarded as an illusion, instead the advantages in giving the evaluator a more involved role are stressed:

> "( )the investigator in action research gets himself involved ( ). Because of the impossibility of appreciating what is going on in human affairs without such involvement, as a matter of deliberate policy action, researchers do not seek the detachment and neutrality that are supposedly necessary for effective experimental and survey work. They argue that such complete neutrality and objective detachment are illusory anyway( )."
> (Orpen 1983, p. 3, quoted in Holt 1985)

What should the Evaluation be used for?

Thorpe (1988) distinguishes between different kinds of evaluation: a *developmental approach* which stresses the importance of evaluation to a continuous development of teaching and learning (supported by Thorpe) and on the other hand an *instrumental or control oriented approach*, likely to be supported at a managerial level:

> "The danger of evaluation ( ) is that it is used narrowly as either a tool for personal or institutional justification or as a means of managerial control. If this is to be avoided, practitioners need to develop an alternative model ( ). It takes a model to kick out a model; that is we cannot expect that a developmental approach to evaluation will come about merely by pointing out the shortcomings of a narrowly instrumental or control oriented approach."
> (Thorpe 1988, p. 202)

Thorpe's wishes concerning the role of the evaluation in a development of learning and teaching also reflect to a high degree the wishes of the Danish evaluation cases analysed within distance education.

However, in Denmark, too, the dangers lurking as described by Thorpe are that the instrumental dimension of evaluations gains ground as part of the general reductions in the field of education.

In particular the area of open education has during the past year been subjected to such economic parameters - such as, for instance, participant's fees, the fact that only the teaching dimension of the open education will be subsidized and not the research work on which the teaching is based, and finally detailed annual approval procedures of all courses - so that one could fear that the more traditional form of quantitatively based evaluation will gain access at the expense of the more process oriented and developmental form of evaluation.

It can be concluded that the Danish organization of evaluation within distance education around cases, as described above, for which staff and money have to be applied for every time means that it will be difficult to sustain a possible pressure from above requesting a more quantitative, instrumental evaluation.

Consequently, one of the goals of Danish distance education might be as soon as possible to establish more permanent evaluation teams who, in line, of course, with the above-mentioned evaluation concept, would have to be supplemented from time to time by persons with an inside knowledge of the process being evaluated. In this way advantages would be gained at the strategic as well as the practical level as it would result in the availability of the desired mixture of external and internal evaluators.

References

(Quotations from books not written in English have kindly been translated by interpreter Mona Lynge, Aalborg University.)

Åbent Universitet (1985): *Åben Universitetsundervisning i Danmark*. Indstilling fra Rektorkollegiets udvalg vedrørende Åbent Universitet. Rektorkollegiet, København.

Bååth, J.A. (1979): *Correspondence Education in the Light of a Number of Contemporary Teaching Models*. Malmø: LiberHermods.

Christensen, 3. B. (1990): Teachers and CMC at Jutland Open University: A Case Study. In: Bates, A.W.(ed.): *Media and Technology in European Distance Education*. Open University.

Dirckinck-Holmfeld, L. et al. (1988): *Teori- og metoderapport*. Picnic-News No.2. Aalborg University.

Dirckinck-Holmfeld, L. (1990): *Kommunikation på trods og på tværs*. Projektpædagogik og datamatkonferencer i fjernundervisning. Picnic-News No.9. Aalborg University. (Ph.D. report)

Fox, D. (1983): Personal Theories of Teaching. *Studies in Higher Education*. Vol.7. No.2. pp. 151ff.

Garrison, D.R. (1989): *Understanding Distance Education*. London: Routledge.

Hald, B. (1987): *Jysk Åbent Universitet 1982-1987*. Aarhus: Jutland Open University.

Holmberg, B. (1986): *Growth and Structure of Distance Education*. London: Croom Helm.

Holt, D. M. (1985): *Theoretical Models of Evaluation*. In: Open Campus. No.11. Deakin University, Australia. pp. 3ff.

Lorentsen, A. (1989a): *Presentation and Analysis of 'Project Computer-Aided Distance Teaching'- a research project at Jutland Open University*. Picnic-News No.4. Aalborg University.

Lorentsen, A. (1989b): Evaluation of Computer Conferencing in Open Learning. In: Kaye, A. & Mason, R. (eds.): *Mindweave. Communication, Computers and Distance Education*. pp. 196 ff. Exeter: Pergamon Press.

Lorentsen, A. (1990, in press): *Brug af systemdata i kvalitativ forskning. Analyser af brugen af datakonferencer på en humanistisk basis-uddannelse på Jysk Åbent Universitet*. Picnic-News No.6. Aalborg University.

Lorentsen, A. et al (1988): *Picnic - Project in Communication Networks in Distance Education Curricula. (Project description in short in English.)* Picnic-News No.3. Aalborg University.

Lorentsen, A. & Dirckinck-Holmfeld, L. (1989): Datamatkonferencesystemer i fjernundervisning. In: Danielsen, O. & Lytje, I. (eds.): *Kvalitative metoder i systemudvikling og følgeforskning*. Aarhus: Aarhus Universitetsforlag. pp. 123ff.

Lorentsen, A. & Rasmussen, P. (1989): Toward a Communicative Distance Education Concept. Openness and Distance Education in Danish Higher Education. *ZIFF Papiere 76*. Hagen: FernUniversität.

Orpen, C. (1983): *Action research in management*. School of Management, Deakin University, Victoria.

Patton, M.Q. (1980): *Qualitative Evaluation Methods*. London: Sage.

Peters, O. (1973): *Die didaktische Struktur des Fernunterrichts*. Untersuchungen zu einer industrialisierten Form des Lehrens und Lernens. Weinheim: Beltz

Søderberg, A.-M. & Anker Olsen, E. (1989): *'Alle Tiders Musik' - evaluering af et mediestøttet åbnet universitetsforsøg*. University of Copenhagen.

Thorpe, M. (1988): *Evaluating Open & Distance Learning*. Harlow, U.K.: Longman UK.

Institutional Evaluation of the National Open University of Venezuela (U.N.A) - A Proposal -

Arlette U. De Machado
Institute for Educational Research
Universidad Nacional Abierta (U.N.A),

&

Leopoldo Machado,
Postgraduate Unit
Universidad Nacional Abierta (U.N.A),
Caracas, Venezuela

I. Introduction

This document examines the evaluation process of distance education at U.N.A (Universidad Nacional Abierta - National Open University of Venezuela). Evaluation here refers to the two aspects involved; that is instruction and administration as conducted at U.N.A, which is a higher education institution that offers an efficient, highly productive and low-cost way of covering the growing demand for higher education.

II. Background

In response to the above mentioned growing demand, in Venezuela a record number of students have gained access to higher education through the distance education program developed and implemented by distance educators at U.N.A.

The U.N.A is composed of various sub-systems aimed at achieving the institution's objectives as well as those of the student population. Therefore, it possesses flexible characteristics and a capacity for self regulation. This insures a progressive up-grading of its activities and products. As we shall see later this capacity influences the evaluation process.

U.N.A therefore is oriented by institutional *principles* and *objectives*. These are:

II.1 Institutional Principles
- Democratization
- Educational Innovation
- Individualized Instruction
- Optimization of the Investment
- National Character

II.2 Objectives
- Institutional Objectives
- Functional Objectives
- Operative Objectives

The *principles* of U.N.A may be defined as follows:

Democratization: U.N.A offers opportunities of higher education to students of diverse social strata, in particular, to those who are employed, or have not had the opportunity to pursue higher education in the traditional system.

Educational Innovation: U.N.A possesses structures capable of developing and incorporating innovations tending to optimize the learning process.

Individualized Instruction: The learning system at U.N.A is based upon individualized study, which is relevant to the conditions, needs and aspirations of the students.

Optimization of the Investment: U.N.A must contribute to a significant reduction of the cost per student as well as the cost per graduate.

National Character: U.N.A extends its higher education services to the entire national territory according to the demand for education.

In correlation with its principles, the *objectives* of U.N.A have been classified in three groups:

Institutional Objectives:

- To train human resources for priority areas of socio-economic development.
- To train professionals who will act upon the social system as agents for the qualitative change demanded by the nation.

Functional Objectives:

- To develop, implement, and administrate new strategies in the learning process.
- To carry out research, to evaluate experiences in the new fields of open and distance education.

Operational Objectives:

- To optimize the efficiency of educational investment.
- To cooperate in the attention to the enrollment in formal higher education so that the latter may increase its efficiency and effectivity.

Finally and for the purpose of this paper we define the following terms:

Distance Education: Implies the utilization of media systems, either individual or collective, in order to offer effective instruction, especially where the *direct contact between students and educators does not exist*.

Open Education: Constitutes the removal of academic restrictions and privileges: the elimination or reduction of the barriers between areas of knowledge, careers, institutions, the increasing and enriching of useful activities and experiences to complement the academic educational purposes, and above all, proposes substantial changes in the traditional relationships between the students and educators.

III. The Problem

Evaluation is defined as the act or result of examining and judging the value, quality, significance, amount, degree or condition of something; and the purpose for evaluating such an institution as U.N.A is to estimate quality, relevance,

performance etc. with a view to correct, improve or change as well as to keep track of the academic evolution of the university, to build a reliable data base and devise a set of criteria and adequate mechanisms for institutional support. The set of data systematically gathered shall provide a basis for a range of activities carried out by the different administrative and academic units of the University.

Therefore evaluation should be considered as a process that implies a comparison of the object under evaluation to another similar object used as a standard of comparison whose qualities are well known to evaluators. Its operational aspects present difficulties, because the evaluation of a distance education system, being dynamic, requires permanent monitoring and restructuring. This, in turn, is due to the fact that the very target of the evaluation is undergoing constant changes.

The above mentioned standards of comparison could be either quantitative or qualitative: in both cases they are man-made and, therefore, as evaluation criteria they cannot have universal value; on the other hand these standards define an ideal state, an acceptable or anticipated behaviour, an intended result or goal etc. which, in turn, implies the need for collecting relevant information on both the exact state of the object for evaluation and the criterion (ideal state) to be used for comparison.

The difficulties involved, in both the selection of all necessary information and the comparison itself is something worthy of special attention. At U.N.A it was concluded that the evaluation system has to have a proper identity and sufficient independence so that it may both perceive and demand compliance with the principles and objectives of the University.

In short, it is a system that on the one hand provides an overall view of the university as a whole, paying attention to the present situation, to the evolution and prospects that apply to academic areas and to results so that they can be used for purposes of institutional support and related administrative activities. The evaluators should, then, be members of the academic and administrative staff who have deep understanding and knowledge of the institution - and therefore it was decided that the evaluation would be carried out only by members of the academic and administrative areas of the U.N.A.

Therefore we may conclude that the institutional evaluation of U.N.A possesses the following general characteristics:

It is an institutional self-study performed by staff; the purpose of such study is to evaluate learning, attitudes, attrition, cost-effectiveness, staff effectiveness and course design; the *key elements* are: committee work, standards set by staff discussion; professionalism.

The risks involved are: alienating some members of staff and ignoring the values of outsiders, but it has as pay offs the increase of staff awareness and the sense of responsibility.

The above mentioned committees were formed on a voluntarily basis and there were no restrictions imposed on the staff; so that they were able to choose the area where they would like to work. Two mayor areas were defined - *Instruction* and *Administration*. Each of these areas shall be analyzed in detail by a committee named the Evaluation Committee for Instruction, and Evaluation Committee for Administration. The aim of the committees is spelled out in their names and they will be carrying out a number of activities in connection with the project that it is hoped will be under way in 1991.

The *Instruction Area* includes the following sub-categories: Design and Validation, Teaching-Learning Process and Student Evaluation Process.

Design and Validation Process. The Curriculum Design determines the profiles, pedagogical objectives, and tasks. These curricular specifications are transformed by the Instructional Design into programs suited for learning and "instructional sets" for the student (written media, audio-visuals, etc). These sets are validated on a representative sample of the population before being applied to the courses.

Teaching-Learning Process. The relation between the student, the learning materials and the special teaching personnel, initiates the phase of interaction of teaching-learning. The student - subject and focal point of the program - studies and carries out the actions established for the achievement of the teaching objectives (Readings, Consultations, Experiments, use of Audio-visual media, Study Groups etc).

Student Evaluation Process. The characteristics of the system urge the student towards constant self-evaluation, so that he has ample and varied opportunities to measure his progress and limitations, and of overcoming the latter, before submitting himself to the final examination. Permanent self-evaluation is complemented by the rigorous and controlled evaluations carried out by the

Institution. From this phase, the information produced by the Summative Evaluation is transmitted to the sub-system of Student Evaluation.

The *Administration Area* comprises all ordinary and routine activities of administration, budget control, personnel and services. This area is divided into three units: General Services, Personnel and Budget Control.

General Services. Processes of internal administration with respect to material needs, commercial functions, maintenance cooperations, filing services, reproduction of material and internal communication.

Personnel. Carries out the administrative processes of the different development programs and control of personnel. Comprises the functions of recruitment and selection, registration and control, classification and remuneration.

Budget Control. Register of Expenditures, Comprises the general accounting process and the register of debits against the annual budget.

References

Castetter, W.B. (1979): *The Personnel Function in Educational Administration*. (3rd edition). New York: Mac Millan

Stock Molsaac, M. (1990): Problems Affecting Evaluation in Developing Countries. *Research in Distance Education*, Vol 2, Number 3, pp. 12-16

Universidad Nacional Abierta (1979): *The National Open University of Venezuela Project*. (1st edition). Caracas: U.N.A

Evaluation in University Distance Education

Horst Möhle

&

Hans-Peter Gerhardt

Karl-Marx-Universität Leipzig

Germany (ex-GDR)

Introduction

Cooperation on the evaluation report of the FernUniversität in Hagen has motivated us to analyse the evaluation practice which has been used since 1952 at the universitas litterarum Lipsiensis. At the same time, this contribution to the report provides us with an opportunity of drawing conclusions for the reorganization of university distance education at our own university which do not only result from our own experience. International findings, backed up by research, which have been carefully processed in the report must also be used to this end.

Whilst on a lecture visit to the Open University in Milton Keynes in May 1990, the first-named author of this paper (Möhle & Stein 1990) was also given the opportunity to get to know, at first hand, research results on "Evaluation in Distance Learning" at the Institute of Educational Technology (Woodley & Kirkwood 1987). These appeared to form a suitable basis for the investigation presented here, which comprises three aspects:

(1) selected fundamental findings on evaluation in university distance studies,
(2) evaluation of the systems, the courses of study and the actual courses in university distance studies,
(3) specifics of evaluating distance students from a didactic viewpoint.

Findings from the academic fields of university and technical college pedagogics and from the faculty for agricultural studies are those which are to be chiefly presented.

Re. (1) Selected fundamental findings on evaluation in university distance studies

When considering the university distance student who is at the same time in full-time employment, it is the social demands which represent the basis of every evaluation. By "social demands" we understand the quality expected in university distance studies, that is, academic-didactic level and effectivity and also the standard which the student has to achieve. Therefore it is necessary to determine, assess and judge, that is to evaluate, the distance study system and the courses of study leading to university degree examinations, and the courses of which all that is composed, as well as the achievements of the distance students.

According to the demands of society, then, we consider the demands put on the distance study system, its courses of study (curricula) and its courses, as well as on personality of the distance student and his/her achievements - and in particular on his/her basic professional competence. The academic niveau is of particular importance at university level. By this is meant the present highest possible level of that special branch of study which is to determine the content of the courses of study and the actual courses. The same is valid for the methodics and didactics of distance education. In the centre of the basic professional competence of the graduate is the competence to do independent academic work, which will be taken up in research work.

All of these requirements are criteria and standards for the evaluation. Until now these aims of education - derived from a requirement characterisation - were laid down in the centrally confirmed curricula and teaching programmes. With the beginning of the 1990/91 academic year they have been defined in University and Examination Regulations (Ministerium für Bildung und Wissenschaft: Anordnung über die Ausarbeitung und Bestätigung von Studien- und Prüfungsordnungen für die Aus- und Weiterbildung an Universitäten und Hochschulen vom 5. Juli 1990).

Owing to the complexity of all educational processes, including that of distance education, the evaluation needs to be as extensive and comprehensive as possible. All of its factors should be given due attention. It must, however, often suffice to include only its typical aspects in the evaluation. Characteristic of our distance education were and still are, for instance, the close connections between the employment process of the students and their study process, the cooperation and communication between lectureres (authors of study letters and those in charge of consultations) and students, which was intended to be as effective as possible, and also the balanced relationship between the development of knowledge and of

skills. Less stress - too little, in fact - was laid, for instance, on the personal responsibility of the distance students for conceiving their own personal course of studies, on the students' working autonomously to a great extent with the study material, on as varied a connection as possible between printed material, audio-visual material and computers. They determine, as objective circumstances of the distance study process, the study activities of the students and all their performance.

These at the same time are the chief content aspects for evaluation.

The approach to be used for the evaluation includes investigation, valuation and judgement of the distance study process and its results. In this connection the collection of data for the chosen variables and the analysis thereof is of particular importance; these are then to be summarized in a report on performance.

As a set of devices, methods of sociological investigation shall be used to this purpose, such as the questioning of individuals or of groups by means of interviews or the use of questionnaires. The results achieved by the students in performance tests and examinations are equally important. Thus quantitative and qualitative analyses are combined.

The evaluation can be done either internally or externally. The factories or firms where the distance student work should also be included in the process. Above and beyond the non-recurring evaluation, evaluation which takes place at the regular intervals over a period of several years is gaining in importance.

Evaluation based on national aims should be supplemented by international comparisons.

The conclusions drawn from the results of the evaluation are aimed at improving the distance study system, raising the quality and effectivity of the programmes and courses, stimulating the lecturers to a better performance but, above all, at stimulating the students to an improved performance. These conclusions become effective during the process itself, or else they serve to cause fundamental changes in university distance studies.

Re. (2) Evaluation of systems, courses of study and evaluation of courses in university distance studies

These basic points for evaluation are specifically assimilated into the systems evaluation as well as into the evaluation of courses of study (curricula) and of courses.

According to Keegan (1986) our university distance education is classified as "consultation distance education". After the most recent developments, however, "seminar or study-course distance education" is a more fitting description. This classification, which is characterized by the large amount of personal cooperation and communication between lecturers and students whilst using study material specially printed for self-study, is of course not a judgement of merit. We think, however, that the specifics of the university level, which exist in the socially necessary formation of distance student personalities who work creatively, definitly live up to expectations.

The most close as possible connection of employment and studies - with regard to contents and organization - is also beneficial to the quality. From this stemmed the consequence, too, to concentrate the academic work done alone by the distance student upon tasks and problems concerning his profession. Whilst doing this, factories and firms were made into the field of examination by the universities when working on research projects. These efforts are definitely considered to be adventageous. On the other hand, however, this basic orientation led in practice to a narrowing of the possibilities for studies. Interests and skills which the workers already possessed remained partially ignored.

The development of any personal responsibility of the in-service distance students for the progress of their own studies remained under-developed in the centrally confirmed study-planning, which also led to a rather limited stimulation for them to work on their studies. The distance student of all students is the one who must choose and build up his own personal course of studies from the large amount on courses on offer, in accordance with the demands of his job, of his studies and of his own interests. The negative evaluation which had so far taken place under this aspect is the reason for introducing basic changes during the present reorganzation of university distance education on the basis of the new University and Examination Regulations.

With view to a modern university distance education which meets with the demands that will be made by a future society, one must also critically assess the

fact that, although the main stress was quite rightly laid upon printed study material, surveys - for provisional orientation - on the complex inter-usage of printed and audio-visual material and computers have so far been lacking.

These qualitative evaluations of our distance education system must be supported by quantitative investigations and analyses. In this connection it is important to note that at twenty of altogether 53 universities and colleges of higher education distance studies ran parallel to face-to-face studies in altogether 130 fields of study. The success quotient in university degree courses actually leading to a degree or diploma is about 70%. This leads to the fact that every forth or fifth graduate of a university or college of higher education received his/her degree or diploma via distance education; all facts of positive value. It is also to be welcomed that sociological surveys taking place at regular intervals on "social determinants in distance studies" were started at the Central Institute for University and Higher Education ("Zentralinstitut für Hochschulbildung", Berlin). The first results from these surveys are available (Gebuhr & Stein 1990), and must needs be lead to consequences for reorganization if they undergo differentiated assessment.

The international comparison which is necessary as a supplement to the national systems evaluation was not sought emphatically until recently, when the co-author of this paper, H. Möhle, laid a broad foundation for it in the study letter "Development of University Distance Education in Important Industrial Countries and in Emergent Countries" (Möhle 1990). In addition, the active participation in international comparative evaluationary research projects is of eminent importance to us in the future.

Closely linked to the systems evaluation is the evaluation of the courses of study (curricula) and the courses themselves. Results of the evaluation of the courses of study and of the courses are also included in the systems evaluation.

So far the courses of study (curricula) for face-to-face and for distance students were worked out together, since the aims and main contents were identical. Curricula confirmed by the Ministry were included in both. They had been prepared and discussed on a broad basis by large teams. The didactic-methodic organization specific to distance education, however, was secured by means of teaching hints for those in charge of consultations and by means of study guides for the distance students. The evaluation of the curricula did not seem necessary in such a procedure.

It was not until during the process of the renewal of professional training in the fields of engineering, agriculture and economic sciences in face-to-face as well as in distance studies - which began in the mid eighties - that the procedure was altered, in that after the drafts of the curricula had been drawn up, a trial phase of five, respectively four and a half years followed for the whole course of study. It was a sort of pre-test before the course of study was generally introduced, and the results brought yearly conlusions for changes as to how far content and methods could be realized. The trial run of the course of study in the field of "Accountancy and Statistics" via distance study at the University of Leipzig is a positive example of this procedure (Studienplan für die Fachrichtung "Rechnungsführung und Statistik im Fernstudium": Karl-Marx-Universität Leipzig, Sektion Wirtschaftwissenschaften).

Internationally seen, of course, the evaluation of courses is in the focus of interest. It aims equally at raising both the level and effectivity of the course material, printed as well as audio-visual. Here, one can follow the proposal of Kirkwood & Woodley (1987), to conduct the evaluation whilst the authors are working out the course material as well as while the distance students are actually using it. They speak of "formative and summative evaluation".

The course of action taken by the Central Office for University Distance Studies ("Zentralstelle für das Hochschulfernstudium", Dresden), which is responsible for the development of distance study teaching material and the formative evaluation thereof, is in conformance with English experience. On the basis of the draft curricula, printed study letters, for instance, are written by a group of authors or one author. Experts (specialists on the field in question and distance education practicians) give their critical opinions on the drafts. Then follows editorial work. The author(s) give(s) the final touches before the work goes into print. The co-author of this paper here bases his statements on his own experience gained whilst working in collaboration with the Central Office.

The summative evluation of course material will be of increasing importance to us in the future. In the process of working for the first time with the course material on hand, it is important for us, too, to seek feedback from the distance teachers as well as from the students during the evaluation which accompanies the distance study process.

Evaluation by the distance teachers should not only include their assessment and judgement, it must also reveal any problems which the students came across whilst working with the material.

Despite subjective assessment and judgement of the teaching material by the students it is also of value to document in as much detail as possible their experiences whilst working on their own with the study material, either by using a questionnaire or by giving individual interviews. The questions raised by the English experts should be taken up: is the aim as regards contents reached in the individual study process? Is the orientation on major content - and academic problems assured? Is the course relevant in practice? Are the contents offered in such a way as to stimulate interested and active study? What study problems occur which are felt to be subjectively problematic? What length of time is needed for the individual study phases? These and similar questions lead to the discovery of starting points for the revision of the teaching and study material or to the alteration in the working methods of the distance teachers. In this respect we wish in future to use existing reserves in order to further qualify the courses.

The improvement of the conceived courses of studies and course materials can, of course, also be guaranteed by an exact determination of the achievements of the distance students, by the assessment and judgement of their personalities and accomplishments.

Re. (3) Specifics of the evaluation of distance students from a didactic point of view

The remarks made so far referred to fundamental findings, as well as forms and methods of the evaluation of distance study systems and courses of study, going as far as assessment and analysis of the students' opinion about their studies.

It is, however, of considerable importance for the overall process of the evaluation that the changes which are to be noted in the students as a result of the course of studies should be registered and assessed as reliably as possible. On the whole, then, it is a matter of evaluating in the sense of assessing the students. The things to be assessed are the results that the student actually attains in his/her studies, which are reflected in the actual outcome.

We consider the outcome, together with the aim, content, methods and organization as being the factual factors and the personal factors (teacher, student) as being an important immanent factor of the study process. The following summary makes clear at the same time the close interactions between the factors of the study process and the respective natural and social fundaments which are to be taken into consideration (Fig. 1).

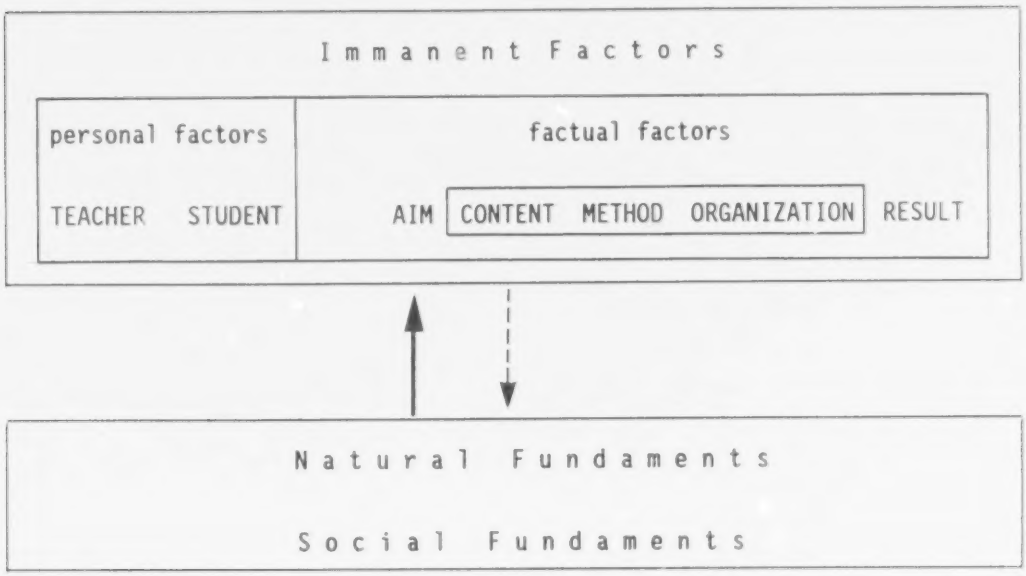


Figure 1: Natural and social fundaments as well as immanent factors of the study process (Autorenkollektiv 1989, p. 15)

One can, then, rightly speak of the result as being an important immanent factor, since with it we view the individual, who has been modelled in the desired way, at a certain stage in his/her development (for us: at the end of a course of distance study); that, then, which is finally decisive from the point of view of the evaluation. This result, the modelled person, is not to be grasped directly but only via his/her image. In the process of constructing this image and of grasping, of establishing results - which is closely connected with the evaluation of results - the function of comparison is of particular interest. It is a question of the comparison between the achieved result and the aim which was originally set for the respective course of study - whereby in the long run the efficiency of the course of study is objectively mirrored. The especially close connections between the factors aim and result (see also Fig. 1) are here again to be seen clearly.

At the same time the question emerges of the structuring of the aims, in order to have at hand manageable criteria for establishing and assessing results. The aim as a mentally anticipated result is oriented, since it aims at modelling the student (personality-modelling), at the structure of the personality. This is shown in the formulation of aims by the indication of qualities pertaining to the personality such as attitude, value-orientation, qualities of will and determination and knowledge, skills and facilities. The more exactly such components of the aims of a course of

study can be given (various steps of the aim; e.g. in how far a person is able to complete an activity reliably and on his/her own), the better they will be able to fulfill their function as objectivized yardsticks within the framework of establishing and assessing results.

It becomes clear that the aim-result comparison, which has here been generally sketched out in all briefness, stands, as an important facet of evaluation in studies, in close relationship to the didactic functional element of control.

As a didactic functional element (by higher deduction, alongside control also guidance, first introduction and consolidation) we regard components of formation of the study process as seen functionally, which are taken up in the immanent factors and realized together with these (Autorenkollektiv 1989, p. 4). This unity of structure and function is stressed at this point in order, also, to be able to despense with having to mention it explicitly in the following expositions.

From this point on we would like to devote our attention towards anwering the question of how establishing and assessing results, in order to discover the actual achievements of the students in their studies during our distance study course (in the field of agricultural studies), was concretely realized.

This on the other hand depends to a large extent on the basic type of distance study in question which, influenced by the historical, national but also geographical (e.g. large territories) conditions, developed in the different countries. The basic types differ according to the dominant didactic regulative, which stands in connection with the feedback; they never occur, then, in a pure form. Likewise the basic types "consultation type", "seminar or study-course type" and "correspondence type". In all of these basic types, however, autonomous study is an integral and essential condition for the success of the distance studies.

Our distance education, as already explained in Section 2, was characterized from the very beginning as chiefly being of the "consultation type". With consultation distance education, too, a further didactic regulative is included, for instance via an introductory course, yearly seminar courses, respectively a final course (usually a one-week course in face-to-face study, taking place at the university/college of higher education), which however, does not act as type-determining. It represents the consultation, which is obligatory and takes place regularly (one day every two weeks or month) at a so-called consultation point, generally in groups of a medium size. This is the main point where teachers and students work together, in between the phases of autonomous study (independent

study) which on its part, as already stressed, is chiefly organized on the basis of mainly written teaching/study materials (Widmann & Gerhardt 1989).

Fig. 2 gives a simplified diagrammatical survey of consultation distance education.

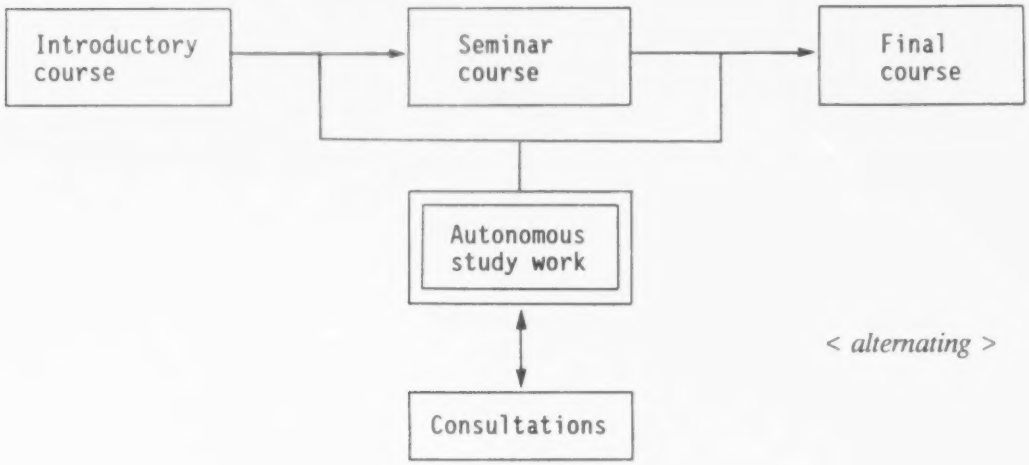


Figure 2: Survey of the course of distance studies (Widmann 1979, p. 17).

The short sketch of consultation distance education made here already shows study phases for which specific possibilities of control must be organized in order to realize the establishment and assessment of results:

- Consultation: The consultation is characterized from a functional viewpoint by the organizational components guidance, first introduction, consolidation and control. However, they present themselves in a weighting and shape specific to distance education. The first introduction should be the exception here, and be put into use - at most - in order to complement or else bring up to date certain contents. Consolidation and guidance (in the sense of introduction into the next phase of autonomous study) have their secured and meaningful place in the consultation due to their position between two phases of autonomous study. What interests us especially is the functional element of control, whereby it is largely connected to the other functional elements and they are reciprocally taken up within each other. Every control, for instance, unconsciously contributes to consolidation, just as the consolidation phase offers possibilities for control for consultation tutors and students alike (cf. Schwarz 1973, p. 19ff).

In consultation in distance study, control serves above all to check the acquisition process which has taken place in the prior autonomous study phase. The consultant draws a picture for him/herself of in how far the student has managed to rise to the demands as regards above all the acquisition of knowledge, skills and facilities. As a rule the control takes place, for this reason, at the beginning of every consultation. Thus the consultant has the possibility of drawing conclusions as to how intense the usually following consolidation phase should be, and what its main focal point should consist of. From this viewpoint it is a matter of controls of achievement in the form of control from the outside. Since every student, though, has the possibility to ask questions during the consultations, or else to bridge gaps in his/her knowledge or to clear up problems which he/she did not understand during the outonomous study phase (form of self-controlling), the person in charge of the consultation must first of all offer these possibilities before any control of the student's performance or assessment of his/her achievement, in the form of mark-giving, can take place.

Tried and tested forms of performance control in the consultations are oral tests (question and answer), giving short accounts, surveying problems, making contributions to discussions, conspectuses, outlining problems, writing short essays and even unscheduled tests ("Klausuren"). Appropriate to the needs of the branch of study in question, programmed performance control was and is also used. In all written performance controls it is of significance to give an assessment in the following consultation, in order to allow the student to take his/her own consequences for his/her further phases of autonomous study.

- Autonomous studies: In the autonomous study phases in consultation distance education, establishment and assessment of results takes place above all as self-controlling and self-assessing performed by the student him/herself, unlike in correspondence distance education. Decisive are, however, also the media via which the teaching personnel indirectly steer the studies and thus also the self-controlling phases.

As already stressed, first of all written material for autonomous study, in the form of special study letters conceived and therefore didactically processed for distance education, was chiefly implemented. These letters containted exercises or control tasks, which allowed the student, by means of solution sections in the appendix and the like, to compare his/her present state of knowledge and skills with the necessary demands, thus enabling him/her to take over the job of controlling the establishment and assessment of results during the phases of autonomous study him/herself. In suitable fields of study (e.g. mathematics and

the sciences) programmed teaching material was also implemented, generally put together as a kind of module. There, the solution of control questions decides whether one can continue on to the next module, or if certain repetition work is necessary.

Courses of distance study on the basis of suitable study letters could usually be absolved with success, especially since the necessary contacts between the authors of the study letters and the heads of consultation could be guaranteed with the support of the Central Office for University Distance Education ("Zentralstelle für das Hochschulfernstudium", Dresden. See also section 2.)

Different conditions under which distance education was organized (for instance the implementation of face-to-face studies and distance studies at a university at the same time), as well as an increasing number of university textbooks and monographs which were conceived for face-to-face studies, led to decisions to implement such teaching and study material in distance study as well. The didactic processing which was here lacking was replaced by specially written study guides, which also included the exercise material and control tests for autonomous studies. Since the heads of consultation were often also the authors of these study guides, a particularly good coordination could be achieved between the self-controlling during the autonomous study phases, which had priority, and the typical control from the outside which took place during the consultations. Also in courses of distance study conducted on this basis, as shown by final examinations and analyses of how successful graduates were ("Absolventenbewährungsanalysen"), the fixed study aim of reaching an identical result in both face-to-face and distance studies could be realized to a very large extent.

- Face-to-face course: Controls of the establishment and assessment of results are also put into use during the seminar courses respectively the final courses presented in Figure 2. The forms implemented there, such as written control of achievements (scheduled written tests: "Klausuren") or special seminars and practical courses in preparation for oral examinations are hardly different in a didactic aspect from the forms implemented in face-to-face studies.

References have already been made to more recent tendencies in development, which, in a part of the universities and colleges of higher education where distance studies are on offer, display a transition to the "seminar or study course type". Although the general appraisal made of the type of course is not

intended to be negated, some problematic aspects linked to this development were, however, to be detected, especially in the field of distance studies in the agricultural sciences:

Due to various reasons (organizational, personnel as well as material), the single consultation days were grouped together to form longer face-to-face periods ("Direktkurse"). During these periods autonomous study was only possible to a certain extent. Between these face-to-face periods, the autonomous study lost much of its original and absolutely necessary continuity, owing to the lengthened phases and the fact that literature suitable for distance education was sometimes non-existent. As many of the teachers in the field of distance agricultural studies confirmed in talks with the co-author of this paper, it came didactically speaking to a kind of "truncated face-to-face studies", where, of course, an identical final standard of both distance and face-to-face studies was no longer to be guaranteed. In possible future developments in distance studies in the field of agriculture, such "operating trouble" ought to be ruled out.

In how far distance studies has a future at all in the said field of studies is to be proved by an analysis of the needs in the course, to be made as part of a continued research undertaking at the University of Leipzig, in which the co-author of this paper will also be involved.

Seen as a whole, a hypothesis of Widmann's (cf. Widmann & Gerhardt 1989, p. 10) might be mentioned, according to which "distance education can take place more dynamically if the basic types are given varied accentuation in order to form the whole course of distance studies. It is assumed that the consultation type is particularly suitable at the beginning of the studies, in order to gradually introduce the distance student to his/her studies", whilst correspondence distance study, which is the basic form making the highest demands on the independence of the distance student, might be stressed more in the later course of studies. The basic types blend with each other, then, during the course of studies, having varying significance, coming close to reaching a kind of optimal level. Investigations need to be made concretely for each special branch of studies as to how this task is to be mastered.

A number of evaluation results introduced in this paper substantiate the said hypothesis.

References

Autorenkollektiv unter Federführung von S. Ulmer (1989): *Wissensspeicher zur Hoch- und Fachschulpädagogik*. Markkleeberg (Hochschulstudium: Agrarpädagogik).

Gebuhr, K. & Stein, R. (1990): Das Fernstudium als Gegenstand bildungssoziologischer Forschung. *Das Hochschulwesen*, 38, 2

Keegan, D. (1986): *The foundations of distance education*. London: Croom Helm

Ministerium für Bildung und Wissenschaft (1990): Anordnung über die Ausarbeitung und Bestätigung von Studien- und Prüfungsordnungen für die Aus- und Weiterbildung an Universitäten und Hochschulen vom 5. Juli 1990

Möhle, H. (1990): *Entwicklung des Hochschulfernstudiums in westlichen Industrieländern und in Entwicklungsländern*: Dresden: Zentralstelle für das Hochschulfernstudium

Möhle, H. & Stein, R. (1990): *Revolutionary renewal of society and the consequences for higher-level distance education*. Milton Keynes: The Open University

Schwarz, R. (1973): *Die Konsultation im Studienprozeß des Hochschulfernstudiums*. Dresden: Zentralstelle für das Hochschulfernstudium.

Studienplan für die Fachrichtung "Rechnungsführung und Statistik im Fernstudium" (Erprobung ...). Leipzig: Karl-Marx-Universität, Sektion Wirtschaftwissenschaften, 1988

Widmann, M. (1979): *Didaktisches Grundschema des landwirtschaftlichen Ingenieurfernstudiums*. Markkleeberg. (Hochschulstudium: Agrarpädagogik)

Widmann, M. & Gerhardt, H.-P. (1989): *Fragmentarische didaktische Skizze zu Fragen der Gestaltung eines modernen Fernstudiums im agrarwissenschaftlichen Bereich (Hoch- und Fachschulen)*. Leipzig: Karl-Marx-Universität (Unpublished)

Woodley, A. & Kirkwood, A. (1987): *Evaluation in distance learning*. Milton Keynes: The Open University, Institute of Educational Technology (IET)

Evaluation Concepts and Methods in Distance Education at Humboldt University (HUB)

Bruno P. Nerlich
Department for Further Education
Humboldt Universität, Berlin
Federal Republic of Germany (Ex-GDR)

1 Development of distance education

The first decree about the establishment of distance education for the working population in the then German Democratic Republic (GDR) was issued in 1950. Since 1951 distance education has been carried out (in the ex-GDR) at the universities of Berlin, Dresden and Leipzig. At first this distance education took place in the field of technical sciences, and has also taken place since 1953 in agricultural sciences and since 1955 in economics and social sciences.

Progress and planning of the contents were determined by the former state department for university and higher education ("Staatssekretariat für das Hoch- und Fachschulwesen"), by means of official enactments and regulations.

The foundation and building up of distance education formed part of an integrated educational system and thereby an integral part of the special tasks of a university or institution of higher education.

The aim of distance education in the fifties was that, running parallel to the face-to-face studies at the universities (which had undergone certain changes), new capacities in education should be created, in order to satisfy the rising need for university graduates ("Hochschulkader") for the economy and administration of the state.

Scientific findings in governmental research policy and methodics and pedagogics, which had been made in the field of adult education at various institutions, were converted for use in university distance eduction.

With the establishment of university distance education as from the year 1950, for the first time in the history of the German educational system larger groups of working people from all productive and non-productive spheres were given the possibility of having a university education without having to take several years off

work. People who had not been allowed to study at a university, be it for political reasons, due to their social conditions or due to the influence of fascism in the Second World War, were now able to do so thanks to distance studies. By the end of 1959 roughly 5,500 working people had received their diploma or degree via distance education (cf. Albrecht et al 1985).

At the end of 1960 there were 31 universities/colleges of higher education, with various branches, offering distance or evening education. The number of graduates had risen within ten years, by the end of 1969, to over 20,000 (cf. Albrecht et al 1985). Until 1975, every third university graduate had got his or her degree via distance studies.

The Humboldt University of Berlin (HUB) takes the lion's share in the distance education of students (in the ex-GDR). Distance education was carried out here, as in other universities/colleges of higher education, in the separate faculties.

Aim, contents and result of the course of studies were the same in both face-to-face and distance education, but due to the character of distance study, the methods and organization of courses had to be different. Lectures, consultations and seminars took place in the faculties, in a fortnightly rhythm for the 40 weeks of the academic year. In the meantime the students used distance study materials and worked on their own.

It was soon to be seen that the composing of distance study units and the execution of lectures for distance students were jobs in which the lecturers had no methodological experience, and jobs which thus proved very demanding. For this reason only the best lecturers could meet with success, and so it was obvious that this new form of education did not meet with the enthusiasm of every lecturer.

Lecturers and distance students, university and the firms in which the students were employed, all worked together closely and via personal contact.

Whilst in the first years it had been the task of distance education to offer instruction to older people who had previously been unable to have a higher education, between 1960 and 1969 the main demand changed. Now the main task was to offer in-service training to younger people in order to prepare them for taking over leading positions later on. Of course, one cannot ignore the fact that the function of distance studies also changed to a certain extent during the seventies.

Although the amount of special subjects offered via distance education increased again in the seventies, the overall number of students actually decreased at the turn of the decade to the eighties (see Table 1). This decrease is, however, not the same in all subjects. Generally speaking, employees in special positions normally requiring a university degree are no longer required to catch up on that degree.

Table 1: Students in distance education at universities and colleges of the ex-GDR

branch of science	1969	1979	1983	1988
mathematics / natural sciences	609	--	--	--
technical sciences	7.579	4.120	3.741	4.090
medicin*	--	--	55	129
agricultural sciences	1.779	996	793	918
economic sciences	8.385	5.195	4.116	4.087
philosophy / state sciences law sciences	3.104	2.033	2.026	3.585
cultural sciences / sports	1.077	775	592	499
literature / languages	274	181	165	193
art sciences	232	371	283	550
pedagogics	8.007	1.089	585	723
Sum	31.046	14.760	12.356	14.774

* incomplete

A number of new laws passed and old laws revised in the early seventies (1971-77) are an expression of the increased demands put upon distance education, defining its position within the continued and further education system and raising the quality of training and eduction, as well as guaranteeing a stable organization.

Since the beginning of the eighties, distance education has been used to prepare people for a future profession or occupation. With this a new social motive

for taking up distance eduction came into being; the wish to prepare oneself for a new phase in one's working life.

This alteration in the social function of distance study meant, at the same time, that the average age of distance students fell; and with this, demans on teaching in distance education rose.

The formation of study groups and collectives of teachers increased the effectivity of studying. The readiness to learn, results achieved and discipline in studying were considered to be good or even very good.

The success quotient in distance education in the ex-GDR lies between 80 and 100 %, apart from in the technical sciences where it is around 50 - 60 %.

At the moment there are 4,600 distance students studying in 25 different subject fields at HUB. Some of these fields are unique in the system of university education in the former GDR (horticulture, the pedagogics of medicine, nursing, crime investigation and food technology, amongst others). Some of the subjects are taught only here and nowhere else. (There is some information about the development of the various subject fields in distance studies at this institution in "Beiträge zur Geschichte der HUB", No. 11, 1985.)

The experiences regarding contents and organization gathered over the decades at HUB have been supported by many analyses. The results of these analyses and their large-scale evaluation helped to raise the niveau of the content and to increase effectivity in distance education here.

We would now like to characterize some of the *methods of analysis* in the following section.

2 Methods for analysing the process of distance education

2.1 Statistical methods

Based on the application forms of the students matriculated in distance studies, the following data have been collected:

- number of students, gender (male/female)
- age, marital status

- social background, level of education
- profession or occupation, function.

On the basis of the curricula, the following points were registered:
- employment of teaching staff, according to whether they were professors, lecturers, assistant lecturers etc.
- days of presence at the university/college of higher educatin, practical courses given in firms/factories or as exam preparation
- examination phases, time for working alone.

The following summaries are based on examination records:
- performance in the particular course of instruction, number of examinations which had to be repeated
- average performance of each registered group of students
- performance in each piece of term work written druing studies, end-of-year papers and degree papers
- number of successful students per year or per registered group of students.

2.2 Direct and indirect questioning

Distance students and lecturers evaluated each semester together, as regarding content and organization. Special emphasis was laid on the assessment of the teaching material which had been used.

Selected registered groups of students and students in specific fields of study received questionnaires, which were to be filled in anonymously and concerned the family and social conditions of the distance students, their field of work and the help which they received from their firms/factories.

For the same registered groups of students the lecturers also received questionnaires concerning the contents and niveau of the lectures and the cooperation shown by the distance students.

The Central Institute for University Education (Zentralinstitut für Hochschulbildung = ZHB) worked out the necessary questionnaire, which was also implemented at other universities/colleges of higher education.

2.3 Evaluation of long-standing experiences

Staff conferences for each discipline and field of study took place once or twice yearly especially for distance education. The experiences gathered there from members of the teaching staff led to quantitative and qualitative statements about the various fields of study and disciplines.

Meetings which took place quarterly with the heads of department in distance studies in each faculty and later section - at HUB there are, at the moment, 14 independent departments - were intended to carefully evaluate the processes of studying in the field of distance education. In these meetings questions of administration and economy were also discussed.

2.4 Separate inquiries

On the basis of a ministry directive, analyses were made of each academic year. For the field of distance education, four or five different points were given chief focus each year, dealing with special questions on the administrative or educative process in distance studies.

Joint activities were made together with the ZHB, within the limits of one of its research programmes, on distance education at HUB. These took the form of direct face-to-face interviewing, questionnaires or else colloquia, problem-oriented seminars, symposia etc.

2.5 Reports - appearing regularly and irregularly - to the head of the university on the distance education of the departments

The results of all analyses were collected and summarized in the Directorate for Further Education ("Direktorat für Weiterbildung") at HUB. Problems were cleared and questions answered or passed on to a higher department, and partially taken over for research.

As examples of the research work in the Directorate for Further Education, comprehensive analyses and evaluations were made on the following topics:

- Inclusion of the distance students into the research work of the departments
- independent academic work done by the distance students

- practice-oriented teaching in agricultural sciences in distance education
- social classification of the students
- ways in which academic information is illustrated
- performance shown by distance students in complex situations
- how the unity of studies and employment is put into practice
- flexible organization of distance study lectures
- aspects of effectivity
- ideas on questions of content and on the organization of the different fields of study in distance education
- proposals for the planning of the respective five-year plans
- efficient administration - central/decentral
- consideration of distance education in new fields of study (making prognoses for education)
- guarantee of supplying and using methods and methodology of academic work
- meeting with the requirements of the curricula by the students, amount of lectures and autonomous study and how these lectures and this study were put into practice.

The sequence of the subjects noted is not meant to show any ranking.

3 Conceptions for distance education

In the early seventies, comprehensive analyses on the development of distance studies at HUB and also at some of the other institutions of higher education in the then GDR were begun. Many discussions and consultations with lecturers and students also took place. All this lead to the first "Conception of the development of distance education and further education at HUB" and in the higher education system of the then GDR altogether. This conception was worked out by the Directorate for Further Education at HUB and ratified by the head of the university in July 1979. This first conception for distance education at HUB served most other institutions of higher education in the ex-GDR as a basis for planning their conceptions of distance and further education. The analyses on higher distance education and the conception of the development thereof, carried out by the Directorate for Further Education, were used at the same time as a HUB contribution to preparing the content of the "5th Conference of Institutions of Higher Education" (V. Hochschulkonferenz) in September 1980.

A continuation and exact specification of the first conception of development followed in 1982. This second conception was also ratified as a "Conception for further and distance eduction at HUB". The aim, as regards content, was concentrated on the tasks of further education, in which distance education is also included, but the chief stress was laid on developing further distance education at the universities/institutions of higher education with a view to making them equivalent in ranking to face-to-face studies and research by 1990.

Following the basic guidelines of a central academic workshop of the ZHB on the theme "The contribution of universities and colleges of higher education of the GDR towards professional further education of graduates from universities and technical colleges", which took place in 1985, and also following directives of the ministry, a third "Conception for the development of further and distance education at HUB for the years 1986 - 1990" was worked out.

This conception, as a continuation of the second one from 1982, has above all as its concrete tasks those which are valid for all academic fields for the reorganization and modification of distance education, on the basis of central decisions.

In order to guarantee an optimal and effective employment of all potentials for distance and further education at HUB, each department of HUB (33 departments in all) had a further education concept of its own as from the academic year 1986/87, based on a ten-point programme.

These 33 conceptions on development were worked out by the departments together with the Directorate for Further Education, coordinated with working partners and ratified by the heads of department and of the university.

The conceptions of the various departments take fully into consideration the tasks of distance education. Depending on the size of the department and the scope of the branch of study, the papers have a length of between 6 and 22 pages.

In June 1988, the 1986 conceptions of each department were checked, and the following main points discussed individually with the heads of department:

- questions and problems of the process of administration in distance education
- classification of distance and further education in the planning process, under consideration of the effective usage of existing funds
- measures of qualification resulting from the research process

- relations with working partners, if necessary, on the basis of arrangements or contracts
- problems encountered in the realization of the conception of further education, considerations as to the further organization of distance education up to the year 1995 and longer.

The result of these deliberations was that all the conceptions were seen to meet with present demands as regards their long-term determination of contents.

A general revision of these conceptions will take place in the very near future when all institutions of higher education have been restructured.

All in all it can be seen that the conceptions of HUB, of the Directorate for Further Education and of the departments served as an important and useful instrument of administration, that they have proved reliable and that they will continue to help towards the qualification of our distance students.

4 Forms of organization of distance education

The five-year distance study course leading to degree-level was so organized in the fifties that between 36 and 52 days' paid release from the place of employment were given yearly, so that the student could take part in face-to-face studies at the universities/colleges. Alongside these days of face-to-face study at the institutions of higher education or their dependencies, work was done on the basis of self-instruction via written study materials. During lectures and seminars at the university/college, the material taught via self-instruction was controlled and guided. Some of the subject matter was also taught directly in the lectures.

At the end of the fifties a combined distance and face-to-face study was created. Here, on additional days, various forms of study were offered for both distance and face-to-face students, taking the form of colloquia, problem-oriented seminars, lectures. By this means it was possible to shorten the distance study courses by at least one year.

As from 1962 it was possible to do a part-time course via distance studies; a little later the possibility of evening studies was added to the offer of full-time studies, with lectures etc. starting after 4 p.m.

A network of outside agencies was built up in the regional capitals ("Bezirksstädte") in the then GDR for the faculties of history, law and the pedagogics of rehabilitation, similar to that for technical sciences. The students came, and still come to HUB, where they are matriculated, for introductory and survey lectures. The rest of the teaching events - in the form of (obligatory) consultations - take place at the outside agencies. In the field of technical sciences, the teaching of students has been carried out in so-called consultation centres ("Konsultationszentren") since 1969. All outside agencies in existence were assigned territorially to the universities or colleges nearest to them, developed into consultation centres and used by the distance students living within that local area.

The first two years of study, after 1975 the first three years (*1*st phase of study) took and take place at these consultation centres. After this the training is put in charge of the university or college where the student is matriculated.

About 600 - 700 distance students who were matriculated at 18 other institutions of higher education were, for instance, given qualifications yearly by HUB between 1971 and 1975. That means altogether well over 2,000 students who took part in studies.

Special courses of study took place on the basis of acknowledged curricula for applicants from special branches of industry, who had similar educational backgrounds and similar functions at their place of work. The study material was sometimes offered in a concentrated form, practica could be absolved at the place of work during normal working hours and special contents were taught.

Women were also trained and qualified for leading positions in their special branch of work as part of study courses offered especially for women.

All special study courses ended with a degree or diploma after 3 to 3 1/2 years. Generally speaking target groups with a special, homogeneous educational background were chosen for these distance studies with a shortened study duration (special distance studies).

Distance study lectures were also used by graduates who were following postgraduate study courses.

Altogether, a high degree of flexibility and effectivity was attained by means of special forms of organization in distance education.

References

Albrecht, K., Herud, G. & Nerlich, B.P. (1985): Zur Entwicklung der Weiterbildung und des Fernstudiums an der Humboldt-Universität zu Berlin. Beiträge zur Geschichte der HUB, Nr. 11, Berlin

Statistisches Jahrbuch der DDR, Staatsverlag Berlin:

a) 1970, 15., p. 311
b) 1980, 25., p. 304
c) 1984, 29., p. 307
d) 1989, 34., p. 316

Evaluation Concepts and Practices at NKI, Norway

Torstein Rekkedal
Research and Development Unit,
NKI, Norway

Introduction

In this article I shall try to describe the concept of evaluation and what evaluation actually means at NKI. In this connection I will touch upon student evaluation. The main emphasis, however, will be placed upon system evaluation - both in micro and macro perspective. It has not been my intention to write a scientific article, but rather to give an account of NKI practices and experiences. I have concentrated on describing concepts, methods and experiences, rather than presenting specific results from research and evaluation projects.

NKI holds the view that evaluation has to be an integral part of and a continuous activity in any educational undertaking. Without evaluation and assessment of students during the learning process, the conditions for learning are far from ideal. In the same way evaluation of the organisation, of the teaching/learning process, of the learning material and media for learning is necessary for institutional learning and important for increasing quality.

20 years ago, as a graduate student at the University of Oslo, I was engaged in the planning of the research project for my degree in education. I was interested in adult education and particularly correspondence education. Till then very little research had been carried out within this area, and my intention was to make a large scale pilot survey on recruitment, study achievements and completions/drop-out rates in correspondence study in Norway. More or less by accident I came into contact with NKI and asked if I would be allowed to carry out my research basing it on NKI files of correspondence students. I was welcomed and given a completely free hand in collecting the amount and type of information I wished for. A year later my thesis was completed.

The information presented did not at all give a glorious picture of the results of the institution. Although the main focus of my research was to find relations between background factors, personal factors and situational factors with different criteria of success in correspondence study, one could not overlook the fact that the completion rates presented were low. The results were, in fact, quite

disappointing, and one could guess that many administrators would have taken the decision to keep such information confidential.

The rest of the story, in my view, gives a picture of NKI's views on institutional evaluation at that time. Instead of being asked to keep a low profile on the publication of results, the adminstration decided that the full report should be distributed to all individuals and organisations involved in or specifically interested in correpondence education in Norway. An abbreviated report containing all factual information was published in Norwegian and English and distributed both nationally and internationally. I remember I was present once when the director was asked how he dared to publish the results of the drop-out part of the study, and his answer was that he did not dare not to publish, since openness of institutional results is necessary for increasing quality. The immediate reaction at the institution was that all employees - academics, administrators and clerical staff - became involved in a project aiming at assessing the organisation and the administrative, teaching and counselling routines and procedures in order to state long- and short-term goals for changes to increase the quality of teaching material and student sevices. At the same time I was engaged to plan and conduct a research and evaluation scheme with the same goal, quality control and quality development of the NKI distance education system.

NKI today

NKI is the second largest non-governmental educational institution in Norway. It is organised as a non-profit foundation with governmental control and quite large financial support from the state. The organisation has four different departments offering services to the public:

NKI College of Engineering is a four year college of engineering training, decentralised and with 18 centres offering full time studies. As opposed to what is happening in many academic institutions today, where one is moving from traditional academic studies into distance education and new modes of teaching, the NKI College was originally developed from an existing distance education programme. The college still differs from most similar institutes in organisation and methods, which may be explained by its inheritance from distance education. The college is the largest engineering college in Norway.

NKI College of Computer Science is a college offering a two year full time course in administrative computer science and information systems development.

It also offers a two-year part time course giving similar qualifications as the first year of the full time programme.

NKI Distance Education is the second largest distance teaching institute in Norway recruiting approximately 10 to 15.000 students a year, or approximately 50.000 course enrolments. The programmes offered cover technical/vocational training, management, computer science and business training on secondary and tertiary level.

NKI Publishing House develops materials both for NKI activities and for sale on the open market.

To support the activities NKI has established a *Research and Development Unit* responsible for research in education and evaluation. The unit has given priority to research and development activities within distance education.

Some elements and dimensions in the concept of evaluation at NKI

Evaluation is a part of most human activities involving decisions, it simply means assessing the quality of something. We evaluate when we buy a suit or even a hamburger, when we decide to go to the theater instead of the cinema, and when we decide which film to see. After the action we assess again to see whether we are satisfied or would make another choice another time.

Systems approach / educational technology

At the general level NKI follows the ideas and principles of educational technology in developing and conveying education. In general models of education and teaching evaluation always plays an important part. Evaluation is the starting point for the feedback loop in the process. At every stage of development the results are in some ways formally or informally, subjectively or objectively assessed and necessary action taken. Ideally, but not as often as we would wish, a total education programme is offered in a pilot version for formal evaluation and revision before the programme is offered on a larger scale. The usual procedure would be to market a new programme in a more or less complete version and carry out successive evaluations, revision work and changes over a period of time.

Single courses and course material within a programme will always be formally evaluated before being offered to the public, and very often presented in preliminary form before "final" production and distribution.

According to the theories of educational technology, evaluation should be a rather objective process, where learning results measured by some clearly defined performance standards are compared with the defined learning objectives (see for instance Popham 1981). Our view is that this might be an ideal situation, at least in some cases. In practice at NKI, however, and I believe in most other distance education institutions, the objectives defined during development of the course are seldom expressed in formal terms suitable for objective control in a strictly goal-related evaluation. The evaluation procedures thus tend to become more subjective than the ideal of the systems approach would prescribe. This seems to be a common observation. Criterion referenced tests including performance standards were originally presented as a promising tool for securing accountability in education, thus constituting a simple and inexpensive tool for decision makers. But in practice over the years the ideal theory has not lived up to expectations. For a discussion of performance standards and criteria, see Burton (1977) and Glass (1977).

Internal or external evaluation?

There has been a lot of discussion in the educational community on who should carry out evaluation to guarantee maximum objectivity and usefulness of the results, see i. e. Scriven (1975). In my view, the question is more acute in situations where external bodies have the responsibility for accrediting institutions and/or programmes according to certain criteria, and in situations where bodies offering grants wish to control the effects of money granted to certain programmes. According to Scriven (1975) there is no easy way around the question of objectivity. There is no way of guaranteeing that external evaluators are not biased, and there is a good chance that external evaluators are not able to go into the institution and programme sufficiently to come up with a useful evaluation design and to get the really important information for the evaluation.

The evaluation may be undertaken by a member of the programme staff, by another member or other members of the responsible unit in the organisation, by another unit within the organisation or by external evaluators contracted by the teaching institution itself or by some external controlling or fund-giving body. In NKI all these different systems exist. I shall give some examples:

External evaluation of the organisation, programmes and courses.

Norway was the first country in the world to regulate distance education by a specific Act, "The Correspondence Scools Act" of 1948. According to this Act all correspondence schools and every single course must be accredited by the Ministry of Education before being put onto the market.

Today the evaluation is administered and carried out by a controlling body, The Correspondence Schools Council.

Evaluation on the institutional level

At regular intervals the Council is supposed to conduct an intensive and inclusive evaluation of NKI's distance teaching activities including marketing and financial systems, tutor and counselling services, staff and teacher qualifications, courses, turn-around time and administrative routines affecting the total quality of the teaching/learning system. A full report is submitted to the school and to the Ministry.

In our view, the institution gains quite some new and valuable insights through these intensive evaluation procedures.

Evaluation of curricula and study programmes

Traditionally, the Correspondence Course Council has not been occupied with evaluation of the specific programmes or composition of courses. The evaluation has concentrated on assessing the whole institution and the individual single course.

However, Norway has a nationwide system of recognition of college and university programmes and exams. If a programme is recognised by the National Coordinating Body, the programme diploma is transferable to national degree programmes. In this connection NKI programmes on tertiary level are evaluated externally. During this process, curricula, textbooks, exams, teachers and examiners are evaluated according to certain criteria defined by the Norwegian Universities.

As mentioned above, The Correspondence Course Council is also responsible for evaluating each separate course. To carry out the evaluation of the course, the Council appoints external subject experts. I believe these procedures may have been justified at the time when they were initiated. Today, however, the system may have more weaknesses than strengths. Some of these are also discussed by Holmberg (1989).

Firstly, to be accepted as objective evaluators by the Council, the experts are recruited from schools and organisations which do not offer distance education. This means that they are very often not fully qualified to evaluate a distance education course. Often they do not perceive the whole notion of distance education, with its specially developed learning material, and the tutoring and counselling procedures which are part of the course as a whole. They will often tend to look at the material in isolation and compare it with the ordinary text books.

Secondly, some may have basically negative attitudes towards distance education as such - the "my subject cannot be taught by distance education" - attitude.

Thirdly, while evaluators appointed from other institutions some years ago could be considered as objective, they might not always be so today. The reason is that external evaluators often are recruited from institutions which, rightly or wrongly, are criticized for not being able to adapt to recent demands from the government or the public to offer less traditional and more flexible studies. In some cases, where student numbers are falling, successful distance courses may be seen as a threat in a more competitive market. Both situations may cause negative attitudes, which may influence the conclusions.

Fourthly, the evaluators very often come to distance education with a teacher/lecture centred rather than a learner/study centred attitude which makes it difficult to understand the pedagogical philosophy behind the "product" they are looking at.

Fifthly, there is always a risk that the external evaluation as such makes the writer and author work more to satisfy the evaluator than the student. In some cases, this problem becomes evident. That is when the external evaluator actually demands changes which the internal course development team based on their own insights into the subject and into distance learning see as a change for the worse.

Internal evaluation

As mentioned above, NKI has set up a separate unit for evaluation and research activities. While external evaluation does have some positive effects on the quality of the teaching system, the function of external evaluation is mainly summative.

Like most other distance teaching institutes NKI is most concerned with formative evaluation and assessment as a means for revision and development (see Scriven 1981).

Internal evaluation is carried out by the unit for research and development both on its own initiative and under contract from the distance teaching unit. We agree completely with the principles put forward by Scriven (1975) for controlling evaluation bias that results should be fed back not only to those responsible for the programme or evaluation object, but also to their superiors. The normal procedures for internal evaluations carried out by the NKI research and development unit include full internal publication.

Evaluation is also initiated and carried out by the staff who are directly responsible for the administration of a certain course or programme. A programme administrator may develop and send out questionnaires or carry out a telephone interview with students and tutors on a course etc. These evaluations tend to be more informal in design, development of instruments and also in the reporting of the results. From time to time the research unit is asked to give consultations on such projects.

In my experience formative evaluation of courses gives the best results when conducted by an internal specialist unit, not directly involved in the operation of the programme, but working in close cooperation with the programme staff. These evaluations should preferably be supplemented by evaluations done as a natural routine by the staff reponsible for the programme, because everyone engaged in education (or any other activity) should see evaluation as an integral part of their work.

It seems clear that evaluation activities initiated and carried out by the programme administration staff should be highly stimulated. They certainly give useful information. At the same time there is no doubt the results, interpretations and conclusions are often biased. The results should be, and often are, discussed

with people less involved and preferably trained in educational research and evaluation techniques.

Sometimes there are good reasons for contracting external support in evaluation. It might be that it is of special importance to be absolutely certain that the results are as unbiased as possible and that they should be accepted as unbiased by the external market. Sometimes we need extra manpower or additional competence.

What or who is to be evaluated?

In the introduction to this article I stated that I would concentrate on the evaluation of teaching systems, not student evaluation. However, as relevant student learning is the ultimate criterion for measuring teaching quality, student evaluation is in fact central in systems evaluation.

Evaluating students

At NKI student evaluation takes place continuously during the studies. Distance students submit their assignments and receive from their tutor comments, corrections and evaluations as subjective overall comments and a formal grade. Courses below tertiary level normally have quite a high submission density, and every submission is expected to cover 5 to 10 hours of study. Tertiary level courses have a far lower submission density, one submission is normally expected to cover 3 to 4 days of full time study. Most courses end with a formal examination, either internal with internal examiners, internal with external examiners from state colleges/universities (following the requirements for national accreditation of college and university exams) or public nationwide exams.

To some extent examination results are used in course evaluation. In our experience feedback from tutors and students, in forms other than simply grades given for individual assignments, gives a better basis for assessing the teaching quality of a course or parts of a course. From time to time, however, comparisons are made between examination results among distance students and full time or part time students taking the same courses. Such comparisons may at least give some indication of the quality of the distance study material.

In general, we are not in favour of putting too much emphasis on comparative studies. One has to take into consideration that students are self selected into the

different forms of study, and that in general, part time students drop out more frequently than full time students. The raw results from some of our statistics indicate, however, that distance students achieve better results than part time on-campus students (Paulsen & Rekkedal 1990).

Objects of systems evaluation

The whole organisation

As described above, the organisation is assessed by external bodies as part of the Ministry's general accreditation procedures for distance education. It seems clear, however, that it is the internal assessment activities which secure necessary development according to external and internal requirements.

Within the field of distance education NKI has carried out extensive survey research programmes to see whether the organisation functions according to its own goals and objectives.

Recruitment and drop out surveys have been considered to be of great importance. Through recruitment surveys we obtain information to decide whether NKI recruits students according to national and institutional policy, so that NKI should function as a national supplement to other institutions in securing possibilities for education despite a person's geographical, social or educational background.

We have carried out drop out surveys to decide whether drop out rates are acceptable. The studies have been designed to find reasons for dropping out and relations between drop out and personal, situational or course related factors. On the institutional level NKI has set specific goals to increase completion rates. On the institutional level, for example, we have found that the average level of course completions is a little above 50%. A three-year plan has been implemented to raise the average completion rates to a minimum of 60%.

When I chaired the Research Committee of the International Council for Distance Education (ICDE) some critisism was raised against the ICDE Research Committee's support of research projects on drop out and completion rates. The basis for the critisism was that the preoccupation of distance educators with drop-out problems lead to the traditional educational community seeing drop-out as a special problem for distance education. At NKI we consider drop-out statistics to be one important indication of institutional success (although we agree that not every drop-out necessarily proves there is a problem in the system). We do not

believe that distance education generally has a larger drop-out problem than other forms of part time adult education (and as far as possible comparisons go, this belief has been confirmed statistically). As distance educators we ought to be proud of accounting for our failures, too, and thus, be able to do something about them.

In another dimension NKI carries out institutional evaluation every year in connection with recurrent long-term planning. During the planning activities an overall assessment of today's situation is undertaken and compared with the situation desired in three years time, and specific long- and short-term objectives for improvement are specified.

Curricula and programmes, courses and media

As mentioned, curricula and programmes are externally evaluated in connection with nationwide accreditation schemes. However, if our aim really is to strive for higher quality, we have to take internal formative evaluation seriously.

When launching new programmes, at least some evaluation activities are always initiated on the programme level. At present an intensive and continous evaluation programme is taking place in connection with the two first presentations of a new programme for managers in the field of health and social services. An external evaluator has been contracted to follow up students and teachers over a period of a whole year. The evaluator is given a free hand in deciding on evaluation design, procedures and instruments. The evaluator has decided to base the approach on Stake's theoretical views on *"responsive evaluation"* (Stake 1976). The case is presented in some more detail below.
Printed material and two-way communication with a distance tutor is the basis of the majority of NKI distance study courses. Thus, high quality of the printed material is a sine qua non if the total system is to function satisfactorily. During the development of the courses, evaluation and assessment take place in different ways. Here are some examples:

An "instructional designer" may be a member of the course development team and go through the material during the authoring phase acting as some kind of "surrogate student" to assess the learning quality and suggest changes.

Preliminary material is scrutinized by external subject consultants to check the academic quality. Before printing the material is controlled by the consultant from the State authorities.

In some cases the course is printed in a preliminary version. With the help of postal questionnaires or more intensive methods, such as telephone or direct interviews, student experiences are collected before final production of the course. The distance tutors are asked to inform the administration about student difficulties, printing errors etc. as a matter of routine.

During the last ten years other media, such as telephone, computers (including computer mediated communication), broadcasts or taped audio and video material have been introduced. When starting courses where new media and technology are involved, NKI often follows the programmes with specific research and evaluation projects. Some of these are shortly mentioned as case studies below. Similar procedures are normally followed to assess financial and educational aspects of new procedures for tutoring, student support and administration.

The distance tutors are the main contacts of the students during their studies. The quality of the tutors' work is extremely important for the quality of the system as a whole. For a long time the work of the tutors has been continuously surveyed. We have also established formal systems for evaluationof tutors. All new tutors have to go through a distance course covering important aspects of distance learning and tutoring. Prospective tutors are evaluated during this course to decide whether they possess the desirable qualities of a distance tutor (see e.g. Bååth & Wångdahl 1976).

Evaluation criteria

Different criteria have been used for assessing quality. Some of these are mentioned below.

Recruitment may be one criterion for deciding whether the institution, programme or course lives up to intentions. It is a national policy in Norway that everyone should have equal opportunities for education and that education on all levels shall function as an instrument for democratisation and equal status in society. It is a fact that adult education does not live up to this ideal goal. In many instances adult education actually increases differences by recruiting more students with a high prior level of education. With a formal institutional policy of

supplementing the state school by expanding opportunities for education, recruitment surveys are important. NKI has carried out a number of such surveys.

On the other hand, studies of *persistence* are important in deciding which groups are best served. We also consider measuring *completion* and *drop-out* to be an important criterion for deciding the quality of study material and student services.

Grading, however, has not been found to function as the best evaluation criterion. The reason is that grading is normally subjective, and is often used individually during the continuous assessment to motivate and stimulate the students in their studies. Even as far as exams go, we have to admit that evaluation results are the result of subjective decisions of examiners, and are not based on criteria-referenced measurements. And, as mentioned above, due to self-selection via recruitment and drop-out, comparisons between results in different student groups do not give valid data for assessing relative quality.

To a large degree *subjective information* from students, tutors and experts is used for assessing the quality of evaluation objects; learning material, teaching processes and student services.

Evaluation designs, methods and instruments

At NKI we have to some extent differentiated informally between *educational research*, *research-based evaluation* and *informal evaluation*, projects carried out by personnel without competence in research methodology.

It seems clear, though, that the results of most research projects may be used for evaluation purposes. We have carried out a number of *experimental research* projects on different aspects of distance education. In testing hypotheses and verifying theories, measures of statistical significance are important. In a scientific sense experimental designs and significance tests are the only way to prove relationships of cause and effect. On the other hand, an administrator, needing information for rational decisions, might very well act on the basis of data not statistically significant, as well as on the basis of a variety of other more or less subjective information.

Thus, both research and research-based evaluation as well as informal projects and continuous assessment projects have their place in the total system of

evaluation. We have carried out *experimental research*, *survey research*, *developmental testing* and *case studies* in this connection.

In line with the discussion above both *quantitative* and *qualitative* analyses are applied. According to our experiences it seems clear that in the case of evaluation of a certain product - whether a specific learning package or a complete programme - for revision purposes, intensive methods based on *depth face-to-face interviews* or *interviews given by telephone* are the methods which offer us more useful information than large scale studies based on *postal questionnaires*, which allow more quantitative statistical treatment and analyses.

Some specific projects

NKI has a long tradition in *small-scale experimental research*. We have experimentally looked into different *follow-up systems* (Rekkedal 1972a/1973a), *turn-around time* (Rekkedal 1983), *introductory course* in study techniques, *telephone tutoring* (Rekkedal 1989) and a *personal tutor/counsellor* system (Rekkedal 1985).

Concerning survey research, our first project carried out an intensive *recruitment* and *drop-out analysis* of one year's intake into the distance study programmes (Rekkedal 1972b). This survey was followed up by two other surveys. The first looked in less depth into three years' intake to see whether recruitment and completion trends had changed; the second was in relation to different procedures and student services introduced during the period. Another study compared recruitment to distance study and to full time engineering courses and followed the students through a four-year period, ending with a completion/drop-out survey. The *longitudinal study* of distance students gave insights into differences beetween the students' plans, attitudes and expectations at enrolment and their experiences and attitudes given at the time of completion or cancellation (Rekkedal 1973b,1976, 1978).

Qvist-Eriksen (1979a, 1979b) carried out a large scale evaluation survey of NKI combined courses (distance courses supported by local face-to-face teaching). Questionnaires were distributed to 142 teachers and 739 students. The teachers were asked to evaluate the quality of the materials, general aspects of the programmes, NKI services and procedures, roles of the distance tutor and local tutor etc. Similar aspects were covered by the students questionnaires. The results

were used for revision of the system as such, specific programmes and procedures and course material.

At present, we have two development projects which are followed by intensive and continous evaluation activities. One concerns the development of a new media/method system in distance education, the EKKO *Computer Mediated Communication Project*. The other concerns the development of a new programme in Management and Administration for the Health and Social Services Sector.

In the EKKO project we have developed a *computer conferencing system* specifically designed for distance education. Computer mediated communication changes a lot of well-known variables in distance education (see for instance the discussion about computer conferencing as being a "new (3rd. or 4th.) generation" or "shift of paradigm" in distance education (Lauzon & Moore 1989, Garrison 1986, Holmberg 1990).

During the last three years we have continuously developed the computer and administrative systems and developed and tried out computer-conferencing-based distance learning in different subjects. Information has been collected systematically, both by the project manager and by the research unit. The methods applied have included *continous logging of communication activities*, questionnaires and interviews with "electronic college" students, prospective students, correspondence students (for comparisons) and teachers, and comparison of examination results. All data have been continuously published and reported. A collection of articles and conference papers has been published by Paulsen & Rekkedal (1990).

The following is an example of a systematic evaluation project as part of specific programme development being carried out at the moment. Three years ago NKI started the planning for developing a two-year distance learning programme in "Management and Administration for the Health and Social Services Sector". The programme in a preliminary version was ready for try-out from 1990. During the development phase external experts had been contracted to evaluate the curriculum plans as they were finalised. Each single course was assessed and accredited by the government before being tested on the trial run. The whole programme was submitted to the National Coordinating Body for University Studies, assessed and became formally recognised as being equal to one year's full time study in the national system.

It was decided not to put the programme onto the regular market during the first year. A contract was signed between one major hospital and NKI to carry out one full trial run. Later another trial group was accepted. During the try-out period major aspects of the programme, individual courses, study material, distance and face-to-face teaching/learning processes will be formally evaluated for updating and revision before the programme is offered on the regular market (Koch 1990).

An external evaluator has been engaged to design and carry out the evaluation programme. The evaluator bases the design on Stake's *responsive evaluation* (Stake 1976, 1981) or *transactional model for evaluation* (Peersen 1990).

The evaluation plan focuses on three main areas:

1. *Frames/situational factors*, e. g. study form and content, administration, student expectations and demands, teacher expectation of their contributions to the programme as a whole, the expectations of students' superiors, plans, time schedules, economy etc.

2. The *teaching/learning process*, what actually happens during the learning programme due to the effects of the influences of the variables of frame/situational factors.

3. *Results*, e. g. the students' experiences of whether their own behavior and attitudes in relevant areas has changed, superiors' erperience of whether the students' behavior has changed, students' ability to solve problems etc., as well as subjective experiences of course quality related to expectations and experiences.

The results will be caused by an interaction of frame factor variables and process variables.

The evaluator wants, in line with Stake's theory, to find out what is of value to the students by gathering relevant remarks from the various participants. She stresses that the aim of the evaluation is understanding rather than giving explanations and conveying propositional knowledge.

During the evaluation process, therefore, she uses different methods. Some examples are:

1. Successive student questionnaires on expectations and experiences.
2. Problem solving tests.

3. Questionnaires for tutors
4. Continuous loggings of experiences and occurences by tutors, on-the-job coordinator and programme coordinator.
5. Interviews with the superiors of the students on their experctations and their perceptions of the students problem solving abilities in relevant areas.
6. Observations of face-to-face sessions.
7. Some selected students continuously log their experiences.
8. Final in-depth interviews with students and superiors related to previous responses.

The evaluation programme is meant to function as *formative* both for the trial period students and for revision of the total preliminary programme.

Conclusion

In this article I have tried to account for some NKI concepts of system evaluation, some practices and experiences. I have not said much about results. I have found it difficult to go into details of results because I find the different surveys, experimental research projects and evaluation projects cover too many areas and dimensions. Describing specific results, I believe, would go beyond the scopes of this article.

There is one important point which I would like to stress. Systematic collection of data and reporting, in our experience, is the absolute fundamental basis for lasting changes and quality developments. We have seen that good ideas may be tried out and lead to positive change of practices. However, if one does not take the time or have the resources to really examine and report on the effect of specific changes and developments by methods which people trust, changes seem to have a tendency to be very short-lived. When people in the system are actively involved in the developments, and the developments are proven to be effective by accepted methods of evaluation, those developments tend to be followed up - and live.

References

Burton, N. W. (1977): *Performance Standards*. Occasional papers series, 11. Evaluation Center, College of Education, Western Michigan University, Kalamazoo.

Bååth, J. A. & Wångdahl, A. (1976): The tutor as an agent of motivation in correspondence education. *Pedagogical reports 8*. University of Lund, Lund.

Franke-Wikberg, S. & Lundgren, U. P. (eds.) (1981): *Att utvärdera utbilding*. Del 1. Stockholm.

Garrison, D. R. (1986): Multifunction microcomputer enhanced audio teleconferencing: Moving into the third generation of distance education. *Int. J. of Innovative High. Ed.*, 3:1, pp. 26-29.

Glass, G. V. (1977): *Standards and Criteria*. Occasional Papers series, 10. Evaluation Center, College of Education, Western Michigan University, Kalamazoo.

Holmberg, B. (1989): *Theory and Practice of Distance Education*. London, Routledge.

Holmberg, B. (1990): A Paradigm Shift in Distance Education? Mythology in the Making. *ICDE Bulletin*, 22, pp. 51-55.

Koch, K. (1990): *Evaluering av høgskolestudiet i administrasjon og ledelse*. (Internal paper). NKI, Bekkestua.

Lauzon, A. C. & Moore, G. A. B. (1989): A fourth generation distance education system: Integrating computer-assisted learning and computer conferencing. *The Am. J. of Dist. Ed.*, 3:1, pp. 38-49.

Paulsen, M. F. & Rekkedal, T. (1990): *Den elektroniske høgskolen*. EKKO-prosjektet, Del III. Oslo/Nadderud, SEFU/NKI.

Peersen, V. S. (1990): *Evalueringsplan*. Internal document.Bekkestua, NKI.

Popham, W. J. (1981): Målbaserade administrative strategier för omfattande pedagogiska system. *In*: Franke-Wikberg, S. & Lundgren,U. P. (ed.) (1981): *Att utvärdera utbilding*. Del 2. Stockholm.

Qvist-Eriksen, S. (1979a): *Kombinert undervisning*. Delrapport 1. Hvorledes klasselærerne vurderer NKI-skolens kombinerte undervisningstilbud. Stabekk, NKI.

Qvist-Eriksen, S. (1979b): *Kombinert undervisning*. Delrapport 2. Hvorledes elevene vurderer NKI skolens kombinerte undervisningstilbud. Stabekk, NKI.

Rekkedal, T. (1972a/1973a): Systematisk elevoppfølging. NKI, Oslo. See also: *Epistolodidaktika*, 1973:2, pp. 57-63.

Rekkedal, T. (1972b): Correspondence Studies. Recruitment,achievement and discontinuation. NKI, Oslo. See also: *Epistolodidaktika*, 1972:2, pp. 3-38.

Rekkedal, T. (1973b): *Tre årskull brevskoleelever* (English summary: Three years' enrolments to NKI-skolen's correspondence courses), NKI, Oslo.

Rekkedal, T. (1976): *Tekniske studier. Korrespondanseundervisning og klasseundervisning*. NKI, Oslo.

Rekkedal, T. (1978): *Tekniske studier. Korrespondanseundervisning og klasseundervisning*. Delrapport 2. NKI, Stabekk.

Rekkedal, T. (1983): The written assignments in correspondence education. Effects of reducing turn-around time. An experimental study. *Distance education*, 1983:2, pp. 231-251.

Rekkedal, T. (1985): *Introducing the personal tutor/counsellor in distance education*. NKI, Stabekk.

Rekkedal, T. (1989): *The telephone as a medium for instruction and guidance in distance education*. NKI, Bekkestua.

Scriven, M. (1975): *Evaluation bias and its control*. Occasional papers series, 4. Evaluation Center, College of Education, Western Michigan University, Kalamazoo.

Scriven, M. (1981): Fördelar och nackdelar med målfri utvärdering. *In:* Franke-Wikberg, S. & Lundgren, U. P. (ed.): *Att utvärdera utbilding*. Del 2. Stockholm.

Stake, R. E. (1976): *Program evaluation, particularly responsive evaluation*. Occasional papers series, 5. Evaluation Center, College of Education, Western Michigan University, Kalamazoo.

Stake, R. E. (1981): Programutvärdering - besvarande utvärdering. *In*: Franke-Wikberg, S. & Lundgren, U. P. (ed.): *Att utvärdera utbilding*. Del 1. Stockholm.

Evaluation in the Distance Courses Established by the Department of Education of "La Sapienza" University (Rome)

Benedetto Vertecchi
Department of Education
'La Sapienza' University (Rome),
Italy

Foreword

The Department of Education of "La Sapienza" University (Rome) established the first distance course in the academic year 1986-87. It was a postgraduate course about "Methods of School Evaluation" mainly - but not exclusively - addressed to primary and secondary school teachers. As well as for the teachers, the course could be of interest for school managers, public administration officials, and those responsible for vocational and in-service training activities. In fact the public welcomed the proposal very warmly. After the first year more than three hundred people enroled. In the following years other courses were established and the number of the students rapidly increased to 4,000 in the academic year 1989-90. The distance courses available in the academic year 1990-91 are listed below:

a) postgraduate courses (students have to enrol personally at the University "La Sapienza", and they pay the fees fixed by the academic authorities):

- *Metodi della valutazione scolastica* [Methods of School Evaluation], one year
- *Didattica* [didactics], one year
- *Complementi di didattica sperimentale* [Supplementary Studies on Experimental Didactics], six months
- *Complementi di didattica dei mezzi di comunicazione* [Supplementary Studies on the Didactics of the Communication Media], six months
- *Complementi di didattica della storia* [Supplementary Studies on the Didactis of History], six months
- *Complementi di didattica della filosofia* [Supplementary Studies on the Didactis of Philosophy] six months

- *Complementi di didattica della matematica* [Supplementary Studies on the Didactics of Mathematics], six months

- *Complementi di didattica delle letterature straniere* [Supplementary Studies on the Didactics of Foreign Literature], six months

- *Complementi di didattica del testo letterario* [Supplementary Studies on the Didactics of Literary Text], six months

b) courses based on an agreement (the courses are established according to an agreement between the University "La Sapienza" and different organizations interested in offering to a particular public the opportunity of a course of study at a university level).
The agreements in force at the moment are the following:

- agreement with the Ministry of Education of the Italian Republic, Central Office for Vocational Education, for two courses: *Elementi di Didattica* [Elements of Didactics] and *Eviluppe e consolidamento della capacità di comprensione della lettura* [Development and Consolidation of Reading Comprehension Abilities];

- agreement with the Government of Canton Ticino (Switzerland) for the courses of *Metodi della valutazione scolastica* and *Didattica*;

- agreement with the Provincial Administration of Bergamo for the courses of *Metodi della valutazione scolastica*, *Didattica* and *Didattica sperimentale*;

- agreements with several schools for one or more of the courses mentioned above.

The solutions drawn up for running the distance courses are based on a very high level of automatization of the procedures. The resources - both financial and human - available for distance courses were very limited. Therefore, it was necessary to steer the organization in such a direction. Besides, the project was to balance a very high level of automatization with an individualized pattern of educational proposal. Therefore didactic organization had to have a great information flow about the previous cultural and professional experiences of the students and about their study activity during the distance course. For this reason it was immediately clear that assessment had to play a strategic part. Assessment has been interpreted as the activity which allows the information about a determined procedure to be collected in its initial, intermediate and final stages.

This is, briefly, the organization of the postgraduate distance courses:

- each course is subdivided into a certain number of units, according to its duration;
- each unit requires 3-4 weeks of study;
- at the beginning of each period, the students receive a postal packet containing the study and exercise material, two assessment tests and the correspondent optic forms for the answers, a letter which explains what to do;
- in the middle and at the end of each study unit, the students answer the test and send the optic forms to the Department;
- after the input of data through an optical scanner, the processing system writes a personalized correction text for each student. The text is produced as a letter which is immediately sent back to the student. From 1991 it will be possible to communicate with a few samples of students through a telematic system: the students will digitalize their answers on the keyboard of a computer connected with the central data processing system via the telephone network, and will receive the answers in the same way.

The Evaluation of the Context

The purpose of the evaluation activities is to be functional to the needs of the organization of distance courses, as they have already been defined. Therefore it is necessary to analyse the context in which the procedures take place and to identify the variables which contribute to its definition. The context is determined both by the elements which describe the *macrosystem*, and by the elements which refer to the specific *microsystem*. The macrosystem is the background of educational activity: the social structure, the production system, the cultural organization all influence education and are influenced by education. The microsystem determines the attitudes which directly influence the educational demand and the answer of the public to the educational offer.

In the field of distance education, it is possible to say that the macrosystem mainly determines the general characteristics of demand and supply; the microsystem is particularly concerned with the choice of educational strategies. The evaluation activity is functional to its purpose - i.e. the regulation of the educational procedures - only if it is based on continuous and analytic information

about the progression of the procedures, and if the main variables which identify the macro- and the microsystem are constantly pointed out. It is possible, especially for the information concerning the macrosystem, to use data archives already available (such as statistical tables to ascertain the consistence of a determinate part of the population). In other cases it is necessary to get the information through specific research (for instance, when the problem is to know precisely what is the attitude of the adult population, or of a determinate part of it, towards the undertaking of an educational proposal; or to appreciate the effects that determinate changes in the production structure could have on the professional profiles).

Gathering information about the macrosystem constitutes a specific evluation activity. In a structure for distance education, this activity could be carried out by a specialized service. In the experience of the Department this task is simplified because almost all the students are teachers. The information which must be carefully gathered concerns the development of the school system, the need of new professional skills etc.

It becomes obviously more difficult to collect information about the microsystem. In this case it is necessary to have data which could very precisely define the profile of the student to whom the course is addressed. It is important to know which skills the student already has, in which way he/she is used to studying, how much time he/she will probably dedicate to the learning activity, which kind of instrumental means are at his/her disposal, what are the complementary study sources he/she can more easily reach. This is the kind of information which must be evaluated to take the wiser decisions about the educational proposal and organization.

The evaluation of the macro- and microsystem can be considered as a preliminary step towards the actual evaluation activities which will take place during the course. It is obvious that these activities must be accurately planned and organized.

Individual Evaluation and Collective Evaluation

In distance education, as in other educational contexts, the evaluation activity is addressed to each single student and to the students as a whole. Moreover the evaluation must consider the structure, the organization and the management of a course. The evaluation concerning the individual student is based on the

information which defines his/her profile (in its social, cultural and affective aspects), on the data related with his/her educational curriculum, on the measurement of his/her learning results. The evaluation concerning the students as a whole roughly has the same purpose, but its attention is concentrated on pointing out the differences between the subjects composing the considered population, in defining its modal characteristics, in recording the collective behaviour at the different stages of the activities of the course, in measuring the collective results at the end of the course.

Although they fulfill different purposes, the evaluation concerning the individual students and the evaluation concerning the students as a whole are not necessarily based on different information, data or measurements. For instance the measurements resulting from a test may be used either to define the abilities reached by each student or to analyse the distribution of the scores as a tool to describe the whole of the population. In distance education the individual and collective evaluation of the students is linked to a defined educational programme. In fact distance education, more than other educational proposals, emphasizes the role of planning, programming and organizing the procedures: we could say that if traditional education is based on an intelligent artisan approach, distance education has to go further, adopting a more industrial attitude[1)].

Therefore it is necessary to cope with the teaching problems fulfilling two apparently opposite conditions: the maximum of homogeneity in all the basic aspects of the service given to the public and, at the same time, the maximum of individualization of the service to satisfy each single student's needs. It is not particularly difficult to satisfy the first condition, but it is much more challenging to fulfil the second one and the individualization could actually be considered as the main variable of the quality of a distance course. It is quite easy to address to the public an indiscriminate message because, in this case, it is not necessary to solve specific instructional problems: even to print a book or to organize a television program sets up indiscriminate messages. The peculiar characteristic of distance education in comparison with other proposals also based on a cultural communication is the possibility of individualizing the message and making it adequate to the student's individual needs.

Distance education conforms to an educational approach which combines the results of most of the recent eductional research.

---

1) cf. Otto Peters' concept of distance education as an industrialized form of teaching and learning: Peters, O. (1973): *Die didaktische Struktur des Fernunterrichts*. Untersuchungen zu einer industrialisierten Form des Lehrens und Lernens. Tübinger Beiträge zum Fernstudium 7. Weinheim: Beltz. *(- ed.)*

Besides, the opportunities set up by the technology of communication have been widely and usefully adopted in an educational situation in which students are not physically present at a place determined for and qualified to allow, educational activities.

Individual Evaluation

The individual evaluation of the students gives an answer to different demands, according to its temporal placement during the learning process. Therefeore, we shall distinguish between the evaluation which takes place at *initial* stages of the process, the evaluation situated in *intermediate* periods and the evaluation connected with the *final* stages.

a) The *individual initial evaluation* is aimed at defining the profile of each student enrolled on a distance course, at the moment at which he/she begins the study activity. The individual initial evaluation considers in particular:

- each student's cognitive frame at the beginning of the learning activity required by the course. From the point of view of the initial evaluation it is necessary to define the specific entry conditions (learning prerequisites) and to plan different individual solutions to equalize the initial situation. The definition of the initial cognitive frame can be based either on certificates, or on data resulting from objective tests, or on information gathered in different ways (for instance questionnaires, conversation etc.). Collecting initial data has the purpose of starting the process of a *diagnostic evaluation*. From the diagnostic evaluation it is possible to draw a *prognostic evaluation*, i.e. the prevision of the results which can possibly be reached. In the Italian school system, as used to be the case in other European school systems, the diagnostic evaluation is only virtually expressed, by accepting the entrance certification to a determined school level (for instance students need a certificate of compulsory school education in order to enrol at the secondary school[2]). We do not have explicit diagnostic evaluation since the entrance exam to secondary education was abolished in the fifties. Some schools which want to program their activity in a more constructive way base their initial *diagnostic evaluation* on data especially obtained, at the beginning of the school year, by administering batteries of tests to the students. The survey of the results then allows the initial cognitive prerequisites to be assessed and a *prognostic evaluation* to be defined. The prognostic evaluation

2) Compulsory education in Italy takes place at infants'/junior school (5 years) and at middle school (3 years). Secondary education begins after these eight years of compulsory education. (- *ed.*)

must not be considered in a deterministic way as a prediction of a positive or negative result. Besides, it has to be regarded as a useful tool in taking early decisions about the individualization of "remedial" work. The same situation arises in the case of distance education: The diagnostic evaluation is implicit when the only data examined is the certification of the student's previous studies. In fact, it is a considerable advantage to examine specific data obtained at the initial moment and to base diagnostic and prognostic evaluation thereon. In this case the educational decisions would be more reliable and more suitable to the individualization of the teaching activity. For instance before the beginning of the proper course, the students could be sorted out according to abilities and could be trained on the basis of specific propedeutic procedures to get the "remedial" work they actually need;

- the reasons for a student engaging him/herself in a study activity: it is necessary to know why he/she decided to choose between distance education and other opportunities (if we suppose that he/she had more than one opportunity). It is important to analyse the specific aspects of the student's affective attitude towards a distance course, because these aspects could compromise the success of the result. Even if he/she decided to enrol on a distance course, a student could be uncertain about this solution and its specific organization; he/she could fear to be alone in coping with the difficulties; he/she could feel that, socially speaking, there is a negative 'halo' (aura) about distance education as a second-range form of education. It is also very important to know the reasons for the refusal, i.e. to know why a person abstains from an educational opportunity based on distance teaching. This type of information can be obtained through structured questionnaires and, somtimes, through private conversations. In any case the most important goal is to dispose, as soon as possible, of the student's personal file. The file should have two main characteristics: it should be both easily consultable and open to further information concerning the student's study activity. The more suitable solution is to organize an automatized archive in which each student has his/her personal position;

- how much available time each student has for the study activity. Especially if the course is addressed to adult students, who presumably go to work, it is necessary to know what is their job, how much free time they have and which part of this they are willing to use to follow the distance course. This kind of information can also be obtained through questionnaires. In this case one should take care to control the correspondence between what the student has stated at the

beginning of the course and the actual time he/she dedicates to the study during the course.

b) the *individual intermediate evaluation* concerns the way each student proceeds in his/her learning activity. Particularly:

- the intermediate evaluation should be first of all aimed at establishing the progress of each student in acquiring the abilities representing the objectives of the single parts into which the course is divided. This evaluation is required in order to find out if the student needs an instruction supplement to compensate his/her specific weaknesses. In other words, the individual intermediate evaluation of the student represents the condition which transforms the undifferentiated educational proposal into an individualized one. The individual intermediate evaluation is called *formative evaluation* when its purpose is to modify the educational proposal and to make it more adequate to each student's needs. The formative evaluation represents a fundamental function of the educational process, because it allows the information to be used with the purpose of giving a different structure to the process from which the information itself derives. When the evaluation of the learning activity is based on the formative approach, it must be as analytical as possible. In fact we can give the student the help he/she actually needs (compensation of the difficulties in the learning activity), only if the offer of further instruction is addressed in a very specific direction. In other words, it is not enough to realize that a student shows some weakness and is not able to reach the intermediate goals as they have been established during the planning of the course; it is also necessary to know exactly in which objectives the student is failing and, if possible, what are the causes of the failure. In this demanding definition, the intermediate evaluation does not only diagnose the difficulties in the learning activity, but also points out the mistakes made by the student. If the activity does not stop at the level of individualizing the difficulty, but proceeds further with the analysis of the mistakes, it is possible to organize a sharper form of individualization. The main tools for the individual intermediate evaluation are *formative tests*. These tests are structured measuring instruments with preformulated answers. Each item of the test stimulates the student to a performance which reveals the mastering of a determined skill or ability. Each item of a formative test should provide a single piece of information to be used seperately with the purpose of starting up personalized procedures. Therefore the data gathered through a formative test do not have to be analysed in their whole (i.e. considering the scoring of the whole test), but item by item. The answer to each item must be

considered, according to its quality, as an element to insert into *a 'tabula praesentiae'* or *'absentiae'*: the individualized procedures refer precisely to the elements of this second *'tabula'*;

- a second aspect to be considered in the intermediate evaluation is related to the comprehension of the task connected with the suggested learning activity. The student's results during the course are not only due to the adequacy of the teaching materials, but also to the whole of the communication which allows the organization of the activities to be run. If a student does not clearly understand what he/she is supposed to do, it is probable that some difficulties arise which do not originate in the study materials but in some fault in the communication of the task. These difficulties can give rise to demotivation: The student feels inadequate to deal with the learning activities and this negative trend could eventually lead to his/her giving up. To face this section of the evaluation strategy it is useful to record the student's behaviour in order to notice anomalous frequencies of wrong behaviours, misunderstandings, mistakes in following the task;

- the intermediate evaluation also considers the evolution of the affective behaviour towards the course. During the activities it can happen the student's attitude changes. These changes can be related to the object of the study activity, the materials and the implements used during the course development, and to the organization, the quality of the interaction (defined in terms of intensity and effectiveness) between each student and the educational structure which directs the course. The evaluation of the affective attitude can be based on periodical questionnaires - if the course lasts for a long time, on observations to be recorded when the student personally contacts the educational structure, on the information which can be gathered in the course of informal communication (such as telephone calls, letters, etc.).

c) the *individual final evaluation* concerns the single students when they conclude the whole of the procedures of a determined course. Particularly:

- the object of the final evaluation is the level of each student's learning results. In determined cases (i.e. when a course lasts for a long period or is composed of parallel segments) the final evaluation does not concern only the gathering of data which takes place at the end of the course, but also the measurements at significant moments of the development of the course (after a group of units, at the conclusion of a determined segment, even if other segments continue or activity related to new segments is beginning). The final evaluation can have the

sole purpose of measuring the level of the student's achievement, or can be interpreted, in a more demanding strategy, as a tool to stimulate the student's performance to assess his/her capacity in gathering and integrating the elements of knowledge acquired during the course. In this second case the final evaluation has the characteristics of a *summative evaluation*. In distance education, when the intermediate evaluation has been planned according to the logic of formative evaluation (i.e. as a preliminary step to activate individualized procedures of compensation), the final evaluation is a summative one. The final evaluation can be based on different technical solutions, not necessarily alternative: it could be a written exam with a non-structured answer; a written semistructured test; an objective test. It is also possible, and in many cases convenient, to adopt a mixed solution. However, the data gathered through the objective tests (such as multiple choice, matching etc.) are more reliable for a summative evaluation;

- it is also important to evaluate the abilities acquired by each student in distance study. It is reasonable to expect that a student who has satisfactorily followed a distance course has also acquired some abilities in organizing his/her study activity autonomously and is able to understand, interprete and translate in operative behaviour the tasks required and stated in the explanation accompanying the study materials. In other words, if it is possible to find out that the student has acquired a specific ability in studying at a distance, it means that the public is developing a positive attitude towards this educational strategy and that the number of prospective students is increasing. From the gathering of descriptive data on the student's behaviour at the different stages of the course, interesting elements for appreciating the capacity acquired by students as distance learners can ensue;

- the final affective attitude is also to be considered in the final evaluation. This attitude is related either to the contents of the course or to its specific organization. Therefore it is convenient to ask the students who have completed a course to fill in a questionnaire. The questions on the questionnaire should find out if the student's expectations have been satisfied and if he/she would enrol on a distance course again etc.

The collective evaluation of the students

The object of the collective evaluation results from the different aspects of the students' behaviour. The collective evaluation takes place in initial, intermediate

and final stages. In distance education, as in other educational contexts, it is necessary for the evaluation activity to emphasize the students as a whole. The purpose of the evaluation is to point out the differences between the subjects composing the considered group, to determine its modal characteristics, to record the collective behaviour at the different stages of the development of the course and to gather the data on the final learning level. Educational research has emphasized the role of collective evaluation as a necessary condition for overcoming the artisan logic of the procedures in educational activities when they are intuitively settled by the single teacher. This approach is still widely represented in face-to-face educational situations, also because it allows teachers to be employed who have modest - sometimes insignificant - specific educational skills. In other words, in the logic of the artisan approach, the quality of the instruction is based on the teachers' qualities (such as intuition and sensibility) considered as personal innate gifts. Even when some teachers obtain good quality results thanks to their "innate" gifts, it is most improbable that high standard levels could arise from educational strategies based on such a philosophy. We must also consider the trend of the last decades: the access to the educational system of all levels of the population has gradually complicated the instruction tasks. This is the origin of many problems which, at other stages of school development, have been solved by simplifying the situation from the social point of view. In other words, it was not the educational proposal which should adapt to the students' characteristics, but the students who had to adapt to the educational proposal. The students who did not have the required characteristics were simply rejected out of the system. The exclusion was a total one, when the student could not even have access to the initial stage of instruction; it was an exclusion *in itinere* when the adequation of the characteristics of the population to the desired standard took place at other stages. The idea of instruction as a fundamental collective right and the consequent development of schooling have gradually undermined the system of social simplification of the instruction task. In this situation the educational activities need to be supported by increasing professional skills. Therefore, a certain amount of satisfying results has been no longer considered to be a sufficient reason to legitimate determined educational procedures. In fact it is not possible to reach the qualitative level required by rejecting the subjects without the right characteristics. In this new context, the qualitative level of instruction is particularly defined by its capacity of offering the students the actual educational proposal they need to reach determined goals. The analysis of the variability in the distribution of the results then becomes a central datum for evaluating the quality of the instruction. This updated philosophy about the general educational system is even more appropriate when we consider the specific field of distance education. In distance education the physical gap between the teacher and the learner hinders

the action of simplifying elements which are possible in a face-to-face situation. Therefore in distance education the qualitative standard is high only if the whole of the procedures have been carefully planned and carried out and if the activities are controlled by a continuous evaluation in order to introduce any necessary modification. In other words, distance education transmutes the artisan approach into an industrial one: The quality of instruction does not derive any longer from the single teacher's ability, but from correct working of the structure which plans, carries out and runs the educational strategies. The collective evaluation of the students can be based on the same data, measurements and information gathered for individual evaluation. For instance the measurement of a profit test can be used either to estimate the level of each student's ability, or to analyse the distribution of the same data among the students as a whole. It is also convenient to consider the collective evaluation in its initial, intermediate and final stages.

a) The *collective initial evaluation* is emphasized during the planning and the carrying out of a course or at its beginning. Its purpose is to collect information, to gather data, to make measurements concerning the behaviour of the prospective and actual population of the course. The collective initial evaluation gives an answer to the needs mentioned below:

- it collects accurate, descriptive references to the cognitive initial characteristics of the students. These references are functional to the evaluation of the students' entry prerequisites, i.e. the repertoire of the skills and abilities they have at the beginning of a new learning activity. In other words, the purpose of the collective initial evaluation is to set a *diagnostic evaluation*. The evaluation of the prerequisites also gives the elements for a *prognostic evaluation* which allows the possible difficulties in the further stages of the activities to be estimated. *Predictive evaluation* is an extension of prognostic evaluation: comparing the data of the entry group with the data of other previous groups and considering that the educational proposal is constant, it is possible to advance a hypothesis about the final results of the course;

- it allows an immediate differentiation of the educational proposal when the analysis of the prerequisites has pointed out a lack of balance in the entry prerequisites. In the planning of a course it is possible either to provide for uniform beginning of the activities or to prepare introductory segments for the students who need them. In distance education the students are not physically present in the educational situation and the study materials have been planned and carried out in advance. Therefore the collective initial evaluation can not be done on the actual students who will enrol in the course, but on a sample

with the same characteristics. For instance in the planning of a course in statistics it will be convenient for educational research to define with accuracy the level of the initial abilities. If the course is addressed to teachers, it will make a difference if they are teachers of the sciences or the humanities. In the first case, the initial evaluation will point out that the level of the prerequisites is higher than in the second case. Therefore, if the course is addressed to teachers of scientific subjects, the first unit will already have a specific content, but if the students are teachers of humanistic subjects, it will be necessary to define an introductory unit to recall abilities probably forgotten and to facilitate the approach to the content of the course. There could be a third case: part of the students are teachers of the sciences and part are teachers of the humanities. In this case the characteristics of the group are not homogeneous and it will be convenient to provide for different entrance levels in order to ensure a further study activity which is as homogeneous as possible;

- in the planning and carrying out of distance courses, it is important to know the motivation of a part of the prospective public to undertake such a study activity. It is also important to know what are the reasons for the refusal of the educational opportunities based on distance strategies. Betweeen the two more definite attitudes - complete acceptance and complete refusal - there is a great number of intermediate opinions, which it is convenient to analyse, particularly when they concern students who have accepted the distance education proposal but are not completly convinced. This aspect of collective evaluation can give useful information to start initiatives to reinforce motivation: the students have to know that they are not isolated as autodidactics merely because they study at a distance. They must consider that in fact they enrol in an educational structure, even if it is not a traditional one, due to the physical distance between teacher and learner;

- although the organization of the course is, in many aspects, flexible in order to adapt to the characteristics of the public, it also requires the scanning of the activities: the organization needs information about the attitude towards the learning activity (of course modal tendencies will be considered). The organization of the courses has to know how much free time the students have and how much of it they can dedicate to the study. The temporal scanning planned in the project of the course can be considered as the best one, but in many cases, especially if a high variablity in these data should arise, it is convenient that the time scheduled in the project could be dilated in order to satisfy the students' needs.

b) The *collective intermediate evaluation* is based on information, data, measurement concerning the collective behaviour of the students during the development of the procedures of a distance course. Particularly:

- the evaluation must be first of all addressed towards establishing the progress in matching up the goals of the study strategies. Related the students as a whole, the intermediate evaluation gathers the data concerning the adequacy of the single parts of the course, the quota of success and failure of the students, the adequacy of the temporal scanning of the activities etc. The elements gathered through the intermediate evaluation can either be used immediately to improve the course (it is possible to introduce timely modification of the aspects which do not seem adequately able to lead to the intermediate objectives which have been established), or recorded and used if it should be necessary to rectify the organization of the course materials;

- a second aspect of the intermediate evaluation concerns the comprehension of the tasks connected with the suggested learning activity. The students' results during the course are not only due to the adequacy of the teaching materials, but also to the whole of the communication which allows the organization of the activities to be run. If the students do not clearly understand what they are supposed to do, it is probable that some difficulties arise that do not originate in the study materials but in some fault in the communication of the tasks;

- the intermediate evaluation also considers the evolution of the affective behaviour towards the course. During the activities it can happen that the students' attitude changes. These changes can be related to the object of the study activity, the materials and the implements used during the course development, and to the organization and the quality of the interaction (defined in terms of intensity and effectiveness).

c) The *collective final evaluation* concerns the results reached by the students of a course at the end of the procedures. Particularly:

- the main object of the final evaluation is the level of the students' learning results. The elements related to the level of the final results allow a *summary evaluation* to be formulated, i.e. an evaluation of the whole of the results. As distance education often uses intermediate evaluation in order to offer the students the possibility of alternative individualized strategies (*formative evaluation*), the final evaluation can be considered a summative one. The

definitions *formative evaluation* and *summative evaluation* correspond, according to Scriven, to particular characteristics of the educational strategy;

- it is also important to evaluate the abilities acquired by the students in distance study. In other words, a different aspect which must be evaluated is the developing, in the target public of the course, of the skills necessary to follow further proposals of distance education. About this aspect of the final evaluation, it is necessary to assess either the students' capacity in organizing the study activity autonomously (when it is required), or the capacity of understanding, interpreting and translating, in operative behaviour, the tasks required and stated in the explanation accompanying the study materials. In other words, if it is possible to find out that the students have acquired a specific ability in studying at a distance, it means that the public is developing a positive attitude towards this educational strategy and that the number of the prospective students is increasing.

- the final affective attitude is also to be considered in the final evaluation. This attitude is related either to the contents of the course, or to its specific organization.

Evaluation at the British Open University

Alan Woodley
Institute of Educational Technology (IET),
The Open University, U.K.

1 Introduction

Evaluation in the sense of "the systematic investigation of the worth or merit of some object" has several meanings within education. Its most common purpose is to grade and classify students, to sort out the successes from the failures. However, here we are concerned with "Programme Evaluation", where the focus is on how well a particular educational programme, curriculum or teaching method works, how it might be improved and how it compares with alternatives.

This type of evaluation has and continues to be done by formal agencies, such as the Inspectorate in Britain. However, during the late 1950's and early 1960's the pace of innovation increased with the introduction of new curricula (e.g. "New Maths" and "New Science"), new teaching methods, and new programmes aimed at disadvantaged groups. The cost and scale of these innovations led to a demand for more evaluation and this has produced an increase in professional evaluators. Today educational evaluation is a legitimate field of intellectual endeavour and has its own associations, journals, conferences and theorists. However, such evaluation tended to be restricted to looking at the effects of particular curricula or teaching styles. With the creation of new distance teaching universities came the need for the evaluation of institutions, both as complete entities and their constituent parts. This paper concentrates on the British Open University (OU) and examines the origins, the evolution and the current status of "institutional research" - i.e. evaluative research that is carried out by and for the institution itself as a means of self-improvement or, as I will go on to argue, of self-defence.

A large part of the evaluative research is carried out by members of the Institute of Educational Technology which currently has over fifty academic staff. However, not all of its members are involved in evaluation and much research that could be categorised as evaluation is in fact carried out by staff in other areas such as the Planning Office, the Business Development and Marketing Office, the Faculties and the Regions. In the next section of this paper I have attempted to

map out the types of evaluation that have or are being carried out anywhere in the University. In later sections I return to the questions of how and why the research agendas and styles have changed over time.

2 Types of Institutional Evaluation

I. System evaluation

a) Basic measures of activity

Any system evaluation must begin with certain basic measures. How many courses have been produced? How many students are there? Hoch many applicants had to be turned away? This data is drawn from administrative records and is presented regularly, often in the form of an annual report or, in the case of the OU, a "Statistical Digest".

b) Measures of efficiency

Allied to measures of activity come those of efficiency. How many students successfully complete the courses? What workload do they attempt? What is the throughput of students? Again this data comes from administrative records and is produced as part of regular monitoring procedures (McIntosh, Woodley & Morrison 1980).

Evaluation of this type has often moved beyond the descriptive to the examination of patterns and causes. For example, *postal surveys* have been carried out asking students why they dropped out of courses. However, these have tended to produce low response rates and answers of dubious validity. A more systematic approach involved a combination of methods including detailed statistical analysis, an understanding of the subjective process of "dropping-out" and an awareness of the different policy options and their likely impact (Woodley & Parlett 1983).

Other efficiency measures have centred on the question of *cost-effectiveness*. On the face of it distance education is a cheap teaching method, but just how cheap or whether it is cheap at all is still in dispute (Wagner 1977 and Mace 1978). Disagreements among the economists centre upon difficulties in making comparisons with conventional institutions. How does one allow for the fact that distance students remain economically active while studying? Is it important that mature students have less years of economic activity in which to employ their new knowledge?

c) Outcomes

In the Open University the measurement of whether adequate learning has taken place has usually been left to the formal *exams* and *assessment system*. However, as noted later in this paper, there have been some attempts to measure the development of distance students as learners using study inventories and in-depth interviews over a period of several years. Also, in one isolated case, Open University economics students were compared with conventional students by administering a *standardised test* of economic knowledge (Lumsden & Scott 1980).

On some OU courses there are no formal exams and *follow up surveys* have to be carried out to see whether there have been appropriate changes in behaviour or attitudes. For example, in one study people who had bought a "Study Pack" on energy efficiency in the home were contacted several months later to see whether they had carried out the energy-saving measures that had been specifically recommended. In another study young mothers from deprived inner-city areas in Glasgow who had taken short "Community" courses were followed up to see what impact the courses had had on their lives (Farnes 1988).

Mail surveys of Open University *graduates* have been carried out to measure the long-term outcomes for individuals following an extended period of OU study (Swift 1982 and Woodley 1988). These studies have looked at personal, occupational and educational outcomes. It has allowed the University to say to what extent its qualifications have been accepted by other educational institutions for admission to post-graduate programmes and by professional bodies for membership. It has also recorded career changes and determined which, in the opinion of the graduates, occured as a result of ther studies. On a personal level, individuals have been asked about other changes such as growth in self-confidence and ability to communicate.

In recent years more attention has been paid to the increasing proportion of students who gain a number of credits from the OU but leave before graduating. It is important to know whether these "dormant" students are leaving because the OU system is failing them or because they have already gained the benefits that they wanted. It is already known that many of these students have used their credits to transfer into full-time courses in conventional institutions of higher education.

The recognition of OU qualifications has also been approached from the other direction. A *survey of employers* was carried out to establish the

standing of the Open University degree and the acceptability of its graduates (McIntosh & Rigg 1979).

Distance teaching can have other outputs besides the more obvious ones and several of these have been looked at. These include the use, acknowledged or otherwise, of OU teaching materials on courses in other institutions (Moss 1979; Glaister & Carr 1986); the passing on of OU materials to other learners (Stainton-Rogers 1984); and effects on the educational motivation of the children of distance learners (Fenster 1982).

d) Programme aims

The Open University is committed to greater "open-ness" and to an increase in social equity. Consequently much of the system evaluation work has been devoted to investigating how far these goals have been achieved. Particular attention has been focussed on formerly *disadvantaged groups* such as women, ethnic minorities, working class people and those with low educational qualifications. As well as measuring their representation among the student body and looking at the progress they make on the courses, barriers to greater participation have also been examined. This has taken the form of surveys of the general public to determine the levels of awareness and knowledge of the institution (Swift 1980), surveys of people who sent for details but decide not to apply (Woodly & McIntosh 1977), and *surveys of applicants* who decline the offer of a place (Woodley 1983).

e) Policy evaluation

Formative evaluation in the policy area has sometimes taken the form of *market research*. Surveys of prospective students and employers have been carried out to measure the likely demand for possible new courses. Surveys of current students have also been used to sound out opinion on various policy options facing the University. For example, one study tested the reactions of Open University students to the possibility of reduced tutorial provision on higher level courses (Thorpe et al 1986).

Policy Evaluation has also taken the form of *monitoring*. The Open University carries out regular surveys to monitor the financial impact of study on its students, thus gauging the effects of fee increases, changes in local authority assistance, the effects of its own financial assistance schemes, etc. (Blacklock 1982). Other survey data on the ownership of televisions, cassette recorders, home computers, etc., can assist course planning (Grundin 1983).

Research has also been used to evaluate the *impact of policy changes*. In one study researchers looked at the effects of an OU policy to "de-register" undergraduates who had made no progress with their studies over a number of years (Heron et al 1986). The results showed that one unforeseen consequence of this policy was that the University had de-registered many of its own graduates.

Finally, policy evaluation studies have also taken the form of experiments or pilot schemes. Perhaps the best known example involved the admission of school-leavers to the Open University. The University and the Government of the day disagreed about the suitability of the OU for school-leavers and so limited numbers were admitted on a trial basis and the outcome evaluated over several years before making a final decision (Woodley & McIntosh 1980).

f) Organisational evaluation

The Open University, just like any other large and complex organisation, can and has been evaluated in terms of its internal arrangements and procedures. In general terms this has involved *scrutinising the financial management* and general "organisation and methods" of the University. More specifically it has involved tasks such as the *monitoring of tutors' marking patterns* and the *turn-around time* for assignments. Evaluations have also been conducted into the course team approach to distance course writing.

II. Course evaluation

The second major strand of institutional research at the OU is "course evaluation", the aim of which is to improve the quality and effectiveness of the teaching and learning that takes place. The evaluation of distance education teaching materials may seek to provide information that can be used during the process of developing or preparing materials or learning experiences - *formative* evaluation procedures, or information about how well the 'finished' instruction has worked in normal use - *summative* evaluation procedures (Scriven 1967). In practice, it is often impossible to draw such a clear distinction, but it provides a useful way of considering methods of evaluation.

Formative evaluation

a) Critical commenting

The great majority of OU teaching materials are prepared by course teams, and *peer review* of draft materials is common place. At an informal level this may simply involve one or more colleagues reading, listening to or looking at draft materials and providing comments in terms of the suitability of content and the style of presentation. On the other hand arrangements may be made for systematic *critical commenting*, with teachers or writers reviewing the materials prepared by all the others working on the same course or programme. Here there is the potential to improve not only individual teaching materials, but also the overall course of instruction. The reactions of colleagues can also be augmented by adopting the more formal procedure of inviting one or more experts in the field to act as Assessors to comment on the draft materials.

b) Developmental testing

Developmental Testing takes place during the preparation phase and involves *trying out draft teaching materials with students*. The feedback obtained is used to guide and inform writers' revisions to the materials before they are committed to print or tape (Nathenson & Henderson 1980). Such testing may range from a fairly informal student try-out of a single piece of teaching, to an elaborate procedure for testing draft materials for a whole course of instruction.

Students study the draft materials in the usual manner and may be asked to undertake any other requirements, for example submitting assignments, attending tutorial sessions, etc. and possibly sitting an examination upon completion of the course. Their comments on and reactions to the teaching can be collected by means of *questionnaires* and/or *interviews* and observations and are fed into the process of revising the course materials for 'final' presentation.

Experience of developmental testing at the Open University (Henderson, et al 1983) indicates the strength of the procedure for the revision of materials within the overall structure of the course and that these can be of benefit to both course writers and students. It is, however, not particularly suitable for enabling major structural changes to be made to the course. In an attempt to allow for greater flexibility, a number of other procedures have been tried that are part formative and part summative (Henderson et al, op. cit.). These

involve collecting feedback from students and tutors on a short-term 'published' version of course materials, to inform revisions to be made for subsequent presentations.

Summative evaluation

The 'product' of course development in distance education is not just the materials that are delivered to students by one means or another. Rather it is the interaction of learners with those materials and other resources, possibly including tutors and fellow students (Thorpe 1979). Summative evaluation procedures are intended to provide information about a course or materials in use.

a) Feedback from tutors

At the OU most of the part-time tutors are not involved in the development of teaching materials for the course(s) they tutor. Their role is to support courses by running tutorial sessions (face-to-face or by means of telecommunications), marking assignments, teaching at summer schools, etc. They clearly have more direct contact with students and mechanisms can be implemented to collect, on a systematic basis, evaluative comments from them on a range of issues. For example they can give their own reactions to the teaching materials, and also accounts of problems their students have encountered in their studies and assignments (Ryan 1982). The experience of tutors in making the course work can provide particularly useful information for subsequent modifications to or adaptations of the teaching materials and instructional arrangements.

b) Feedback from students

Feedback is gathered frequently from OU students while they are taking a course or shortly after its completion. In some cases it may be possible to implement some revisions during the presentation of a course as a result of students' comments, for example by providing a supplement to update information or to clarify a problem area. More frequently, the student feedback from one presentation of a course helps to determine revisions for subsequent presentations. After a course or programme of instruction has been presented in substantially the same form to many cohorts of students, feedback may be collected to inform decisions about remaking or replacing the course. Information gained by course writers about the success (or otherwise) of approaches and strategies employed in their distance teaching may prove to be of great value when they prepare further courses.

Mail questionnaires are the most widely used method for collecting feedback from students and the types of information sought tend to fall into the folloqing areas:

* Extent of utilisation. Students may be asked to indicate which parts of the course or programme they have studied, which components they have used, how much time they have spent on their studies, etc. They may also be asked to report on any problems they have encountered in obtaining the course materials or in gaining access to resources.

* Overall view of the teaching. Students may be asked to rate the teaching of a particular unit of instruction in terms of its interest, perceived relevance or usefulness, level of difficulty, etc. They may also be asked to rate individual components of a course (e.g. teaching text, audio-tape, etc.) in terms of their relative usefulness.

* General style of presentation. Course writers may be keen to receive students' comments on the style of presentation, both in terms of layout, design, etc. and the coherence and clarity of the teaching. Perhaps more importantly, students could be asked to comment on the extent to which the teaching style or strategy had enabled them to become actively engaged in learning from the materials.

* Specific Content Issues. It is important to know how well the teaching has achieved its aims and objectives. To this end, information about students' problems with key concepts, ideas and relationships, etc. can be of great value to course writers when it is time for revisions to be made.

Cross-sectional studies

Some evaluation work has involved *study of a particular innovation or component* used in a number of courses. The aim of such studies has been to draw out generalisations from the use of a particular aspect of the teaching, or to establish the effectiveness of a particular strategy or teaching medium.

The role of particular course components has been of concern to the Open University and has been the focus of much research. Research into the use of *audio-visual media* (Grundin 1985), *tutorials* (Kelly 1981; Kelly & Swift 1983), and *computer assisted learning* (Scanlon et al 1982) are examples of this type of work.

The research has involved collecting information from students on their access to and use of particular components and resources, as well as eliciting their views on the contribution made to the teaching and the overall effectiveness of courses.

The introduction of an innovatory teaching strategy has also given rise to evaluative research across a range of courses, for example the use of project work to encourage greater independence in distance education (Henry 1979).

Developmental studies

Some evaluation of teaching effectiveness has taken the form of *research into the understandings* developed by students as a result of their studies. Qualitative changes in the understanding of key concepts and relationships formed the focus of a study of Open University social science students (Taylor et al 1981a, 1981b), based upon a method for evaluating the content of students' learning developed in Sweden by Dahlgren (1978).

The study was concerned with finding out not *how much* students know, but *what* they understand about particular concepts and principles. A group of university students taking an introductory social science course were individually asked a set of questions about key concepts and principles taught in the course. They were interviewed before commencing their studies and again after completion of the course. The findings of this study informed the writers of the replacement course, not only by identifying problem areas in the teaching, but also by illustrating the different levels of pre-course understandings that students were likely to have.

3 The Evaluation Agenda

When the Open University was created (it received its Royal Charter in 1969 and the first students began their courses in 1971) it was always intended that its activities should be closely monitored and evaluated. One can adduce several reasons for this:

- The University had to be accountable to the Government which was the main source of funding.
- The University was to be funded directly by the Department of Education and Science rather than through the Universities Grants Council and therefore

would be scrutinised as a unique institution rather than as just another university.

- The University represented a major innovation in terms of teaching methods.
- The teaching material would be available for public examination and would therefore need to be of a consistently high standard.
- The students were to be admitted on a "first-come, first-served" basis, regardless of prior educational qualifications. It would be necessary to know just who the students were and what progress they made.

Whatever the reasons, formative evaluation activities were already underway in 1970. Some of this took the form of developmental testing with volunteers being recruited to try out the brand new foundation courses. Other researchers were studying those people who had enrolled on preparatory courses designed by the National Extension College, their purpose being to get an idea of who the first OU students were likely to be. The staff involved in this work were soon brought together in the new and rapidly expanding Institute of Educational Technology. Most of the types of evaluation listed in the previous section have been carried out over the years by members of the Institute and the impression gained by the reader might be of a comprehensive, carefully integrated evaluation programme. However, the programme has been patchy and its development has been affected as much by external and internal pressures as by considered debate.

In his review of access issues in the OU, West devised a typology which distinguished between the concepts of "Market", "Liberal" and "Radical" access policies (West 1989). One can attempt to characterise different evaluation practices in a similar fashion.

"Market evaluation"

Here evaluation corresponds to the sort of research done by a profit-maximising manufacturing company. Students are seen as customers and research is devoted to maximising their number and their throughput.

"Liberal evaluation"

With this type of evaluation the researcher is the students' friend. Information is gathered on why people drop-out, what they think is wrong with the courses, etc., so that the system can be made better for them.

"Radical evaluation"

This can take several forms but essentially involves the evaluator in taking a critical stance concerning which people become students, what they are taught, how they are taught and the effects on their lives.

Each of these evaluation strands exists within the OU and, somewhat confusingly, the same research study can be operating in all three modes. Nevertheless it is worth trying to document how the balance has shifted between the three over the years and to consider the causes.

"*Market evaluation*" was of key importance in the early years when the OU's viability was in question. Was it attracting sufficient students and was it producing graduates in a cost-effective fashion? This has been a continuing concern but takes on greater importance in years when the flow of new applicants slows down or when, as in 1990, the Department of Education and Science mounts a review to see how well the OU is performing. As mentioned in the introduction, this is when evaluation clearly takes the form of self-defence rather than self-improvement. Here the role of the institutional researcher is to provide the information that the University can use to convince its paymasters that the high level of state-funding is justified. In the present debate this includes demonstrating that the OU produces large numbers of graduates in Maths, Science and Technology, that its graduates perform well in the labour market, and that it attracts large numers of students from educationally disadvantaged groups.

Market evaluation has grown in importance outside the undergraduate programme. In the Continuing Education Programme the courses have to be self-financing and so particular attention has been paid to how many students a course is likely to attract and whether their employers would sponsor them. Members of IET have been involved in such work but it has increasingly come under the remit of the Business Development and Marketing Office. This office in turn has frequently used outside market research companies to carry out surveys whose aim is to find ways of maximising sales.

"*Liberal evaluation*" has almost been synonymous with course feedback. Providing such feedback is always listed as one of IET's primary tasks but it is surprising how little effort has been devoted to it in recent years. In 1971 large numbers of students completed detailed questionnaires on each unit of each foundation course throughout the year. In 1990 the main source of feedback comes from the Annual Survey of New Courses. This is a single questionnaire completed by samples of students after the courses have finished and, as the title implies, only

affects courses in their first year of presentation. The lack of attention paid to this area has several causes:

- The great incrase in the number of courses being offered.
- The decrease in staff numbers in IET.
- The routine nature of this work means many researchers find it boring and there are no junior research staff to delegate it to.
- The type of information collected does not often lead to simple messages for course improvement and the nature of OU course production means that very little actually can be changed in any case.
- Academics who have been in the OU for a long time feel that they know how to produce good courses.

Developmental testing has virtually disappeared, mainly for the same reasons but also because of the cost and the difficulty of finding appropriate "students". In recent years much attention has been given to courses that involve "home computing". The studies that have been carried out tend to focus on the practical difficulties involved in acquiring the equipment, paying for it, installing it, und using it in the home.

"*Radical evaluation*" has always existed on some level but is enjoying something of a resurgence at the moment. For example, following a Senate motion re-committing the University to the goals of widening access and of equal opportunities, much attention has been paid to how many people from various disadvantaged groups such as ethnic minorities and the working class have entered the OU, what courses they take and what progress they make as students (EO 1990). Various pilot schemes are being carried out and evaluated in the regions to see how things can be improved.

IET is also about to undertake a *longitudinal study* of a sample of OU students. This will involve re-interviewing students over a number of years in order to study the processes of learning with the OU and how this is integrated with their everyday lives (Morgan & Farnes 1990).

Critical evaluation of the curriculum has taken place in the context of the Equal Opportunities Report. Recommendations have been made concerning the content and style of all courses, the provision of courses that focus on issues of discrimination and disadvantage, and appropriate staff recruitment and training. Elsewhere the critique has focussed on the ideological bias of the curriculum and the modes of delivery (Burt 1989, 1990).

4 The Future

Under the present Government the OU is being pressured to be less reliant on state funds and to generate more income from student fees. Faced with a shortage of 18 year-olds due to a demographic dip, other institutions are going out of their way to recruit mature students. These two forces are likely to lead to an even greater emphasis on market evaluation. If IET staff are not willing to do it then it will be done by other units such as Buisiness Development and Marketing Office (BDMO) or the Planning Office or it will be bought in from outside.

Given their status as tenured academics, researchers in IET will be able to pursue radical evaluation as long as they continue to receive the support of the OU's academic community. However, they will almost certainly be required to do so with less financial funding and, as the first generation of staff begin to leave the OU, they will almost certainly not be replaced by evaluators with the same security and freedom to determine the evaluation agenda.

Acknowledgement

This paper grew out of an earlier one written with Adrian Kirkwood for the Commonwealth of Learning:

Woodley, A. & Kirkwood A. (1988): Evaluation in Distance Learning. Student Research Centre Report No. 21, IET

References

Blacklock, S. (1982): What the OU cost the undergraduate student in 1981 (mimeo). *Survey Research Department Paper No. 229*, The Open University.

Burg, G. (1989): Computers in schools as culture and ideology. *ICDE Bulletin*, vol 20, 20-25.

Burt, G. (1990): Ethics and management education. *ICDE Bulletin*, vol 24, 27-34.

Dahlgren, L.O. (1978): Qualitative differences in conceptions of basic principles in economics: A contribution to the discussion of the validity of examination results. *4th International Conference on Higher Education*, University of Lancaster.

EO (1990): *The Report and Statistical Digest of The Equal Opportunities Team* (mimeo). The Open University.

Farnes, N. (1988): Open University community education: emancipation or domestication? *Open Learning*, Vol 3, No 1, 35-40.

Fenster, E. (1982): *College attendance by working adults and its effects on the educational motivations of their children*. Detroit: To Educate the People Consortium.

Glaister, B. & Carr, R. (1986): Education open to all. *Open Learning*, Vol 1, No 3, 50-52.

Grundin, H. (1983): Audio-visual media in the Open University (mimeo). *IET Papers on Broadcasting No 224*, The Open University.

Grundin, H. (1985): *Report on the 1984 audio visual media survey* (mimeo). Institute of Educational Technology, The Open University.

Henderson, E. et al (1983): Developmental testing for credit: A symposium. *Teaching at a Distance Institutional Research Review*, 2, 39-59.

Henry, J. (1978): *The project report* (mimeo). Institute of Educational Technology, The Open University.

Heron, M., Kelly, P. & Marshall, J. (1986): Student progress and non-progress (mimeo). *Regional Research and Development Papers*, No 15, The Open University.

Kelly, P. (1981): *An overview of student use and appreciation of tuition* (mimeo). RTS Research Group, The Open University.

Kelly, P. & Swift, B. (1983): *Tuition at Post Foundation Level in the Open University - student attitudes towards tuition* (mimeo). RAS / IET, The Open University.

Lumsden, K. & Scott, A. (1980): *An output comparison of Open University and conventional university students* (mimeo). Edinburgh: Esmee Fairbairn Research Centre, Heriot-Watt University.

McIntosh, N. & Rigg, M. (1979): Employers and the Open University (mimeo). *Survey Research Department Paper No 162*, The Open University.

McIntosh, N., Woodley, A. & Morrison, V. (1980): Student demand and progress at the Open University - the first eight years. *Distance Education*, Vol 1, 37-60.

Mace, J. (1978): Mythology in the making: Is the Open University really cost-effective? *Higher Eduction*, Vol 7.

Morgan, A. (1984): A report on qualitative methodologies in research in distance eduction. *Distance Education*, Vol 5, 252-267.

Morgan, A. & Farnes, N. (1990, in press): New directions in research and evaluation: Longitudinal studies of student learning. *Student Research Centre Report No 40*. IET, The Open University.

Moss, C.D. (1979): The influence of Open University distance teaching in higher education. *Teaching at a Distance* 14, 14-18.

Nathenson, M. & Henderson, E. (1980): *Using student feedback to improve learning materials* London: Croom Helm.

Ryan, S. (1982): Developing tutor feedback. *Teaching at a Distance*, 22, 9-14.

Scanlon, E. et al (1982): Computer assisted learning. *Teaching at a Distance Institutional Research Review*, 1, 59-79.

Stainton-Rogers, W. (1984): Alternative uses of materials: research findings. *Teaching at a Distance*, 25, 58-68.

Swift, B. (1980): Trends in awareness and beliefs about the Open University among the general public (mimeo). *Survey Research Department Paper No. 185*, The Open University.

Swift, B. (1982): What Open University graduates have done (mimeo). *Survey Research Department Paper No. 230*, The Open University.

Taylor, E., Gibbs, G. & Morgan, A. (1981a): The outcomes of learning from the Social Science Foundation Course: Students' understandings of price control, power and oligopoly (mimeo). *Study Methods Group Report No. 9*, The Open University.

Taylor, E., Gibbs, G. & Morgan, . (1981b): Students' understandings of the concept of social class (mimeo). *Study Methods Group Report No. 10*, The Open University.

Thorpe, M. (1979): When is a course not a course? *Teaching at a Distance*, 16, 13-18.

Thorpe, M. et al (1986): Effective study in the Open University: The human dimension (mimeo). *Student Research Centre Report No. 1*, IET, The Open University.

Wagner, L. (1977): The economics of the Open University revisited. *Higher Education*, Vol. 6.

West, B. (1989): *The UDACE/RAS/R04 Report. Developing access*: A case study of the Open University, with special reference to the West Midlands region (mimeo). The Open University.

Woodley, A. (1983): Why they declined the offer. *Teaching at a Distance*, 23, 2-7.

Woodley, A. & McIntosh, N. (1977): People who decide not to apply to the Open University. *Teaching at a Distance*, 9, 18-26.

Woodley, A. & McIntosh, N. (1980): *The door stood open: An evaluation of the Open University younger students pilot scheme*. Lewes: The Falmer Press.

Woodley, A. & Parlett, M. (1983): Student drop-out. *Teaching at a Distance*, 24, 2-23.

Woodley, A. (1988): Graduation and beyond. *Open Learning*, Vol 3, No 1, 13-17.

List of Contributors/Addresses

Jesusita Alvarado B.
Jefe del Centro de Control de
Calidad Académica (CONCAL)
Universidad Estatal a Distancia (UNED)
Sabanilla, San José,
Costa Rica
FAX-No: 56 534990

Guiselle Belaño M.
Centro de Control de Calidad Académica (CONCAL)
Universidad Estatal a Distancia (UNED)
Sabanilla, San José,
Costa Rica
FAX-No: 56 534990

Guiseppa D'Agostino S.
Jefe del Centro de Control de Calidad Académica (CONCAL)
Universidad Estatal a Distancia (UNED)
Sabanilla, San José,
Costa Rica
FAX-No: 56 534990

Dr. Jörn Bartels
Zentrum für Fernstudienentwicklung (ZFE)
FernUniversität
Postf. 940
D-5800 Hagen 1 (W)
Federal Republic of Germany
FAX-No: 49 2331 688896

Angela Castro
Head of Academic Programs
School of Continuing Education
Hong Kong Baptist College
224 Waterloo Road
(2nd Floor, 322 Junction Road)
Kowloon, Hongkong
FAX-No: 852 3 384040
email address: DCEangela@bc750.Bitnet

Dr. Margalit Ganor
Department of Evaluation and Didactic Training
The Open University of Israel (OUI)
No. 16 Klausner St.
P.O.B. 39328
Ramat Aviv 61392
Israel
FAX-No: 972 3 6423639

Dr. Hans-Peter Gerhardt
Agrarwissenschaftliche Fakultät
Wissenschaftsbereich Landwirtschaftl.
Hoch- und Fachschulpädagogik
Karl-Marx-Universität
Dölitzer Str. 8
D 7113 Markkleeberg (O)
Federal Republic of Germany

Dr. David Granger
Director, Center for Distance Learning (CDL)
EMPIRE STATE COLLEGE
State University of New York
Saratoga Springs, NY 12866-4309
U S A

Prof. B.N. Koul,
Director, Distance Education
Indira Ghandi National Open University
Post Bag 18, Hauz Khas P.O.
New Dehli-110 016
India
FAX-No: 91 011 650863

Dr. Timothy Lehmann,
Associate Vice President
Office of Resaerch & Evaluation
Empire State College
State University of New York
Saratoga Springs, NY 12866-4309
U S A

Annette Lorentsen, Associate Professor
University of Aalborg
TNP, Teknologi og ny
Pedagogisk Institut 12
Aalborg Universitetscenter
P.O.Box 159
DK-9100 Aalborg
Denmark
FAX-No: 45 98 157303

Arlette U. de Machado, PhD, Associate Professor
Institute for Educational Research
Universidad Nacional Abierta (U.N.A)
Av. Los Calvini (antes Av. Gamboa) No. 18
San Bernadino, C a r a c a s 1010/A
Venezuela
FAX-No: 58 25 743086 (Bibliotheca CRM)

Leopoldo Machado, PhD, Associate Professor
Post Graduate Unit
Universidad Nacional Abierta (U.N.A)
Av. Los Calvini (antes Av. Gamboa) No. 18
San Bernadino, C a r a c a s 1010/A
Venezuela
FAX-No: 58 25 743086 (Bibliotheca CRM)

Dr. sci. Horst Möhle, Professor (em.)
Agrarwissenschaftliche Fakultät
Wissenschaftsbereich Landwirtschaftl.
Hoch- und Fachschulpädagogik
Karl-Marx-Universität
Dölitzer Str. 8
D 7113 Markkleeberg (O)
Federal Republic of Germany

Dr.sci. Bruno Nerlich, Direktor
Direktorat für Weiterbildung
Humboldt Universität zu Berlin
Unter den Linden 6 / PSF 1297
D 1086 Berlin (O)
Federal Republic of Germany

Torstein Rekkedal
Director of Research and Development,
N K I
Hans Burumsvei 30
Postboks 111
1341 Bekkestua
Norway
FAX-No: 47 02 530500

Dr. Rudolf Schuemer
Researcher
Zentrales Institut für Fernstudienforschung (ZIFF)
FernUniversität
Postf. 940
D-5800 Hagen 1 (W)
Federal Republic of Germany
FAX-No: 49 2331 880637

Prof. Benedetto Vertecchi
Dipartimento di Scienze dell'Educazione
Corsi di Perfezionamento a Distanza -
Universita di Roma "La Sapienza"
Via del Castro Pretorio, 20
00185 Roma
Italy
FAX-No: 39 6 4452642

Alan Woodley
Research Fellow
Institute of Educational Technology (IET)
The Open University (OU)
Walton Hall
Milton Keynes MK7 6AA
England, U.K.
FAX-No: 44 908 653744